

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71-191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.4%; Score 5918; DB 1; Length 1260;
Best Local Similarity 86.4%; Pred. No. 4,8e-234; Indels 14; Gaps 5;
Matches 1092; Conservative 51; Mismatches 105;

Qy 1 MELAAACRWGLLLALLPGCAASTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60
Db 4 MELAAACRWGLLLALLPGIAGTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 63

Qy 61 ELTYLPTNASLFLQDIQEVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 120
Db 64 ELTYVPANASLFLQDIQEVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 123

Qy 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDF---NNFT 176
Db 124 DPQDVAASTPGRTPEGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDFVFRKN-- 181

Qy 177 VSFMLRVPKVSASHLENRSRACHPCSPMKGSRGWGESSEDCQSITRTVCAGGCARCKGP 236
Db 182 ----QLAPVDID--TNRSRACPPCAPACKONHGWGESPEDCQILITGICTSGCARCKGR 234

Qy 237 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRYT 296
Db 235 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRYT 294

Qy 297 FGASCVTAPNYLSTGVSCITLVCLHNOEVTADGTQRCCKSPCARCYGIGMEHL 356
Db 295 FGASCVTTCPNYLSGVSCITLVCLHNOEVTADGTQRCCKSPCARCYGIGMEHL 354

Qy 357 REVRAVTSANTQEFAGCKKIFGSLAFPEFSDGPDASNTAPLOPQLOVFTELEITGYL 416
Db 355 RGARAITSDNVOEPDGCKKIFGSLAFPEFSDGPDASNTAPLOPQLOVFTELEITGYL 414

Qy 417 YISAWPDSLPDLSPFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLAIHH 476
Db 415 YISAWPDSLRSLSVFQNLRIIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLAIHR 474

Qy 477 NTHLCFVHTVPDQLFRNPHQALLHTANRPEDE-CVGRGLACHOLCARGHCWGPCTOCV 535
Db 475 NAHLCFVHTVPDQLFRNPHQALLHSGNRPEDELCVSSGLVCNSLCAHGCWGPCTOCV 534

Qy 536 NCSQFLRQECVEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADOCVACAH 595
Db 535 NCSHFLRQECVEECRVWKGKLPREYVSDKRCLPCHPECPQNSSETCFGSADQCAACAH 594

Qy 596 KYDPFPCVARGPSGVKPLSYMTWPKFDEGACQPCPINTHSCVDLDDKCPAERQAS 655
Db 595 KYDSSSCVARGPSGVKPLSYMTWPKFDEGACQPCPINTHSCVDLDDKCPAERQAS 654

Qy 656 PLTSTVSAVGLLVVGLVGVFGILIKRQOKIRKYMRLLOETELVELPLTPSGAMPNQ 715
Db 655 PVTFTIATVEGLVFLILVGVFGILIKRQOKIRKYMRLLOETELVELPLTPSGAMPNQ 714

Qy 716 AQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 775
Db 715 AQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 774

Qy 776 EAYMAGVGSYVSRRLGICLTSTVQLVTQMPYGCILLDHVRENRGRIGSODLLNWCQI 835
Db 775 EAYMAGVGSYVSRRLGICLTSTVQLVTQMPYGCILLDHVRENRGRIGSODLLNWCQI 834

Qy 836 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWM 895

835 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWM 894
Qy 896 ALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICIT 955
Db 895 ALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICIT 954

Qy 956 DYMTIMVKCWMIDSCRRPRFRELVSERFARMARDPQRFVVIQNEDLGAPSLDSTFYRSL 1015
Db 955 DYMTIMVKCWMIDSCRRPRFRELVSERFARMARDPQRFVVIQNEDLGAPSLDSTFYRSL 1014

Qy 1016 EDDMDGLVDABEYLVOOGFPCDPAPACAGCMVHRHRSSTRSCGGDLTGLLEPSEEE 1075
Db 1015 EDDMDGLVDABEYLVOOGFPCDPAPACAGCMVHRHRSSTRSCGGDLTGLLEPSEEG 1074

Qy 1076 APRSLAPSEAGSGSDVDFGDLGMAAKGLQSLTHDPSPLQRYSEDPVPLPSETDGYVA 1135
Db 1075 PPSPLAPSEAGSGSDVDFGDLGMAAKGLQSLTHDPSPLQRYSEDPVPLPSETDGYVA 1134

Qy 1136 PLTCSPPQYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVKDVFAGG 1195
Db 1135 PLACSPQYVNPQSEVQPPPLTPGGLPVPVRPAGATLERPKTLPSPGKNGVVKDVFAGG 1194

Qy 1196 AVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDQDPPERCAPSTFKGTATENPEYL 1255
Db 1195 AVENPEYLVPREGTASPPHPPAFSPAFDNLYWDQDSEBQGFPSNFEGTPTAENPEYL 1254

Qy 1256 GLDVPV 1261
Db 1255 GLDVPV 1260

RESULT 3
148161
P:185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Iehika
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:g493236; PIDN:BA003801.1; PID:g747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.4%; Score 5914.5; DB 2; Length 1254;
Best Local Similarity 86.4%; Pred. No. 6.6e-234;
Matches 1092; Conservative 59; Mismatches 100; Indels 13; Gaps 4;

Qy 1 MELAAACRWGLLLALLPGCAASTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60
Db 1 MELAAACRWGLLLALLSPGASGTQVCTGTDMLRLPASPETHDMLRLHYQCCVQVQNL 60

Qy 61 ELTYLPTNASLFLQDIQEVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 120
Db 61 ELTYLPTNASLFLQDIQEVQYVLIHNOVQVPLQRLIRVGTQLFEDNVALAVLNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDF---NNFTV 177
Db 121 DPLNNTTATGRTPEGLRELQRLSLTEILKGVLIORNPOLCYODTILWKDFVFRKN--- 177

Qy 178 SFMLRVPKVSASHLENRSRACHPCSPMKGSRGWGESSEDCQSITRTVCAGGCARCKGP 237
Db 178 ----QLAPVDID--TNRSRACPPCAPACKONHGWGESPEDCQILITGICTTAPRAVPAARL 231

Qy 238 PTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTF 297
Db |||
Qy 232 PTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTF 291
Db |||
Qy 238 GASCVTACPNYLYSTDVGSCTLVCPHNOEVTABDGTQRCCKSPCARVCVGLGMEHLR 357
Db |||
Qy 232 GASCVTTCPNYLYSTEVSCTLVCPHNNOEVTABDGTQRCCKSPCARVCVGLGMEHLR 351
Db |||
Qy 358 EVRAVTSANIQEFAGCKKIFGSLAFLEPSFGDGPASNTAPLOEQLOVPETLEEITGYLY 417
Db |||
Qy 352 GARAITSANIQEFACKKIFGSLAFLEPSFDGNPSGTAPLTPEQLOVPETLEEITGYLY 411
Db |||
Qy 418 ISAWPDSLDLSVFQNLQVIRGRIILHNGAYSLTLOGLGISWGLRSLRSLGSLALIRHN 477
Db |||
Qy 478 THLCFVHTVPWDLFRNPHOALLHTANRDEDCVEGLACHOLCARGHCWGPGPTQVCNC 537
Db |||
Qy 472 THLCFVHTVPWDLFRNPHOALLHSGNPSEECGLKDFACYPLCAHGHGCGPPTQVCNC 531
Db |||
Qy 538 SFLRGQECVEBCRVLQGLPREYVVARHCLPCHPECPONGSVTCFGRPADOCVACAHYK 597
Db |||
Qy 532 SHFLRGQECVCKRWKGLPREYVNGKCLPCHPECPONSTETCTGSEADOCCTACPHYK 591
Db |||
Qy 598 DPPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPL 657
Db |||
Qy 592 DSPFCVACPSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQASPA 651
Db |||
Qy 658 TSIVSAVGVILLVVLGVVFGILLIKRRQKIRKYIMRLLQTELVEPLTPSGAMPNQAK 717
Db |||
Qy 652 TSIIATVGVILLVVLGVVVGILLIKRRQKIRKYIMRLLQTELVEPLTPSGAMPNQAK 711
Db |||
Qy 718 MRILKETLRKVKVLGSGAFGVYAGIWIPOGENVKIPVAIKVLRNTSPKANKELDDEA 777
Db |||
Qy 712 MRILKETLRKVKVLGSGAFGVYAGIWIPOGENVKIPVAIKVLRNTSPKANKELDDEA 771
Db |||
Qy 778 YVMAGVSPYVRLLIGICLTSTVQLVTLQMPYGLLDHVRNRLGSLGSDLLNWCWQIAK 837
Db |||
Qy 772 YVMAGLSPYVRLLIGICLTSTVQLVTLQMPYGLLDHVRNRLGSLGSDLLNWCWQIAK 831
Db |||
Qy 838 GMSYLEDLVRLHRLDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKWMAL 897
Db |||
Qy 832 GMSYLEDLVRLHRLDLAARNVLKSPNHVKITDFGLARLLIDETEHADGGKVPKWMAL 891
Db |||
Qy 898 ESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 957
Db |||
Qy 892 ESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 951
Db |||
Qy 958 YMIMVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLED 1017
Db |||
Qy 952 YMIMVKCWMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLED 1011
Db |||
Qy 1018 DMGDLVDAAEYLVPOQGFCCPDPARGAGMVHHRSSSTSSGSGDLTLGLEPSEEEAP 1077
Db |||
Qy 1012 DMGDLVDAAEYLVPOQGFCCPDPARGAGSTAHRRSSSTSSGSGDLTLGLEPSEEEAP 1071
Db |||
Qy 1078 RSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPVLPSPSTDGWVAPL 1137
Db |||
Qy 1072 RSLAPSEGAGSDVFEGLGMCATKGPQSI SPRDLSPLQRYSEDTPLPTETDGYVAPL 1131
Db |||
Qy 1138 TCSPOPEYVNPQDVPPOPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAFGGAV 1197
Db |||
Qy 1132 ACSPOPEYVNPQEVPRPPLTPEGLPLPPVRPAGATLERPKTLSPGKNGVVKDVFTFGGAV 1191
Db |||
Qy 1198 ENPEYLTPOGGAPOPHPPAPSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLG 1257
Db |||
Qy 1192 ENPEYLVFRGGASQFH-PPALCPAFDNLVYWDQPPSERGSPNTFEGTPTAENPEYLG 1250
Db |||
Qy 1258 DVPV 1261
Db |||
Qy 1251 DVPV 1254
Db |||

RESULT 4
GOHUE
epidermal growth factor receptor precursor - human
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Homo sapiens (man)
C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999
C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143;
R:Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.
J. P.H.
Nature 309, 418-425, 1984
A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression
A:Reference number: A00641; MUID:84219729; PMID:6328312
A:Accession: A00641
A:Molecule type: mRNA
A:Residues: 1-1210 <ULL>
A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
A>Note: The authors translated the codon AAG for residue 540 as Asn
R:Ikhi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A:Title: Characterization and sequence of the promoter region of the human epidermal
A:Reference number: A25772; MUID:85270438; PMID:2991899
A:Accession: A25772
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-29 <ISH>
A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA52370.1; PID:g553272
R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificatio
A:Reference number: S30024; MUID:88217333; PMID:3329716
A:Accession: S30024
A:Molecule type: DNA
A:Residues: 1-29 <HA2>
A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R:Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A:Title: Contributory effects of de Novo transcription and premature transcript termi
A:Reference number: A38672; MUID:91107677; PMID:1988448
A:Accession: A38672
A:Molecule type: DNA
A:Residues: 1-29 <HAL>
A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
A:Experimental source: carcinoma cell line A431-7
R:Xu, Y.; Ikhi, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.;
Nature 309, 806-810, 1984
A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RN
A:Reference number: A00642; MUID:84245835; PMID:6330563
A:Accession: A00642
A:Molecule type: mRNA
A:Residues: 'RCAMRA', 150-187, 'KSVIOAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-
, 798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF re
R:Liu, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma,
Science 224, 843-848, 1984
A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
A:Reference number: A43615; MUID:84196372; PMID:6326261
A:Accession: A43615
A:Molecule type: mRNA
A:Residues: 713-964 <LIN>
A:Experimental source: epidermoid carcinoma cell line A431
R:Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A:Reference number: A23062; MUID:85046483; PMID:6093780
A:Accession: A23062
A:Molecule type: mRNA
A:Residues: 1028-1210 <SIM>
R:Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A:Reference number: A05281; MUID:84172183; PMID:6324343
A:Accession: A05281
A:Molecule type: protein
A:Residues: 25-30, 'S', 32-51; 454-467 <WEB>

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C:Species: Gallus gallus (chicken)
 C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 R:Liaw, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
 A:Reference number: A27720; MUID:88261272; PMID:3260329
 A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 A:Cross-references: GB:M20386
 R:Nilaen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643, MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M10066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 specific protein kinase
 F:1-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-1223/Domain: signal sequence #status predicted <SIG>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:31-654/Domain: EGF receptor extracellular domain repeat <EE1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.4%; Score 3110.5; DB 1; Length 1223;

Best Local Similarity 48.5%; Pred. No. 1.1e-119;

Matches 632; Conservative 172; Mismatches 349; Indels 149; Gaps 25;

QY	8	RMGLLLALLPPCAA-----STQVCTGTDMLKRLPASPEHLDMLRHLYQCGOVQGNLE	61
DB	13	RGAALVLLVLLGVALCSAVEEKVCGTNNKLTQLGHVEDHTSLQRMYNNECVLSNLE	72
QY	62	LTYLPTNASLSFLQDIQEVGYVLIHQNQVQVPLQRLRIVRGTLQFEDNYALAVLDNGD	121
DB	73	ITYVEHNRDLTEFKTIQEVAGVYLIANMVDVPLENLQIRGNVLYDSFALAVLSNVH	132
QY	122	PLNNTTPTVGSPGIGRELQRLSLTEILKGGVLIQVLPOLCVQDTILKDFNFTVFWL	181
DB	133	-MNTQ-----GLRELPMKRLSEILNGGVYKISNNPKLCNMVDVLMNDIIDTS----	178
QY	182	RVPKVSASHLENRSRACHPCSPMKSGRCWGSSESDCSLTETVCAGCA-RCKGLPTD	240
DB	179	RKPLTVLDFASNL-SCKPCHPNCTEDHCWAGEQNCQTLTQVICAQCSGRCKRVPSD	237
QY	241	CCHECAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGAS	300
DB	238	CCHNCAAGCTGPRESDCLACKKFRDDATCKDTCPLVLYNPTTYQMDVNPBGKYSFGAT	297
QY	301	CVTACPNYLTDSGCTLVCPHLNQVETADGTORCKSKPCARVCYGLGMEHLREVR	360
DB	298	CVRECPHNVVYTDHSGSVRSNCNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGI	356
QY	361	AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTEITGYLYISA	420

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A:Title: Ligand-specific activation of HER4/pi80erB4, a fourth member of the epiderm

DB	357	SINATNIDSFKNCTKINGDVSLPVAFLGDFAFTKTLPLDPKKLDVFTVKEISGFLIQA	416
QY	421	WPSLSPLDSVFNQVIRGILHNGAYSILTLOGIGISWLGRLSRLRELGSGLAIHNTHL	480
DB	417	WPDNATDLVAFENLEIRGTRKHGOYSLAVNVLKTSQGLRLSKISDGDIAIMKVNKL	476
QY	481	CPVHTVPDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHGCMGPOTCVNCSQF	540
DB	477	CYADTMNMRSLFATOSQTKIINRNKNKNDADRHYCDPLCSDVGCMGPGPFHCFSCRF	536
QY	541	LRGQECVECRVLQGLPREYVNAHCLPCHPECPONG---SVTCGPEADQCACAHYK	597
DB	537	SROKECVKOCNIIQEPREFEDSKLPCHESECLVQNSTAYNTTCSGPGDCHMKCAFI	596
QY	598	DPFFCVARCPGKVPDLSPYMPKWFDEEGACQPCPINCTHSCVDLDDKGCPAEORASPL	657
DB	597	DGPHCVKACPAVLGENDTL-VWKYADANAVQLCHPNCTRGCKGPGLEGCP---NGSKT	652
QY	658	TSIVSAVV-GILLVVVLGVVFGILIKRRQOKIRKYTMRLQLQETELVEPLTPSGAMPNOA	716
DB	653	PSIAAGVVGGLLCLVVGGLGIGLYLRRR-HIVRKRTLRLQLQERELVEPLTPSGEAPNOA	711
QY	717	QMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDE	776
DB	712	HLRIILKETEFKVKVVLGSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDE	771
QY	777	AVMAGVGPYVSRLLGICLTSTVOLTPMLPVGCLLDHVNRGRGLSGDOLLNWCQVIA	836
DB	772	AVMASVDNPHVCRLLGICLTSTVQITQIMPGYGLLDYIREHKNIGSYLNLNWCQVIA	831
QY	837	KGMSYLEDVRLVHRDLAARNVLVKSFNHVKITDFGLARLLDIDETEHADGGKVPKWWA	896
DB	832	KGMNLEERLVRDLAARNVLVKTQHVKITDFGLAKLGADEKEYHAEGGVPIKWWA	891
QY	897	LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLLEKGERLPPOPICTID	956
DB	892	LESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPPOPICTID	951
QY	957	VYMWKMWIMIDSECRPRELVSFRSMARDPQRFVVIQ-NEDLGASPLDSTFVRSLL	1015
DB	952	VYMWKMWIMIDADSRPKFRLIAEFSGNWARDPPRYLVIOGDERMHLPPTDSKFRVTLM	1011
QY	1016	EDDDGDLVDABEYLVPOQGFPCDPAPGAGGMVHRRHSSSTRSGGDLTLGLEPSEE	1075
DB	1012	EEDMEDIVDADEYLVPHQGF-----NSPST-----	1038
QY	1076	APRSP-----APSEGASDVDFDGLMGAAKGLQSLPHTDPSPLQRYSEDPTVPLPSET	1130
DB	1039	-SRTPLLSLSLTSATNSATNCID-----RNGQGHVPVREDSFVQYSSDPTGNFLEES	1089
QY	1131	--DGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLPKGNGVVK	1188
DB	1090	IDDGFL-----PAPEYVQ--LMPKKPS-----TAMVQVQIYN	1120
QY	1189	DVF-----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQ--	1231
DB	1121	NISLTAISKLPMSRYQNSHSTAVDNPEVL-----NTNOSPLAKTVFESSPYIQSG	1172
QY	1232	-----DPPE-----RGAPPSTFKGTPTAENPEYVLGLDVP	1260
DB	1173	NHQINLNDPDYQQDFLPNETKPNGLLKVPAENPEYLVRAAP	1214

A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLOS>
A:Cross-references: GB:U07868; NID:G337359; PID:AAB59446.1; PID:G337360
A:Note: sequence extracted from NCBI backbone (NCBI:P126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.4%; Score 2972.5; DB 2; Length 1308;
Best Local Similarity 45.1%; Pred. No. 4.7e-114;
Matches 613; Conservative 185; Mismatches 369; Indels 191; Gaps 31;

QY 9 WGLLLALLPPGAA-----STQVCTGTDMLRLPASPETHLDMLRHLXGQCVVQGNLELTY 64
DB 8 WWSVLSLVAAAGTVQPSDSQVACGTENKSSLDLEQQVRAALRYENCCEVVMGNLEITS 67

QY 65 LPTNASLFLQDIQEVQGYVLIHAHQVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLN 124
DB 68 IEHNRDLFLRSRVRETVGYVLVALNQFVYLPLENLRIIRGTLKYEDRYALAIFLNRYKDG 127

QY 125 NTPVTGASGGGLREQLRLSLTEILKGGVLIQNPOLCYODTILWKDF-----NNFTVS 178
DB 128 NF-----GLQELGLKXNLTEILNGVVVDQNKFLCYADTIHWQDIVRNWPSPNLT- 177

QY 179 FWLVRVKNVSHLENSRACHPCSPKCGSRCKGSESDCOSLTRVTCAGGC-ARCKGPL 237
DB 178 -----VST-----NGSSGCGRCHKCTG-RCMGPTENHCQTLRTVCAEQDCGRGYPY 225

QY 238 PTDCHQCAAGCTGPKHSCLACLHFNHSGIGELHCPALVYNTDTFESMPNPEGRYTF 297
DB 226 VSDCHRECAGCGSGPKDTCFACMFNDSGACVTCQPTFFVYNTPTFQLEHFNKAYTY 295

QY 298 GASCVTACPNYLLSTDVSGCTLVCPLNHQBNVTAEDGTQRCCKSKPCARVCYGLGMEHLR 357
DB 286 GAFVCKVKCPHNFV-VDSSSCVRACPSSKMEV-BENGIMCKPKPTDICKACDGIQTGSLM 343

QY 358 EVRAVTSANIOEFAGCKKIFGSLAFILPESFGDPGASNTAPLOEQLOVRETLEITGYLY 417
DB 344 SAQTVDSSNIDKFINCTKINGNLIFLVTHGHPDQYNAIEAIDPEKLNFRVREITGFNL 403

QY 418 ISAWPDLPLDSVFQNLQVIRGILHNGAYSLTLOGIGISWLGSLRLSRLGSLALIHNN 477
DB 404 IQSWPPNMTDFSVFSLNLTIGRVLVSGLSLLIKQGGITSLOFQSLKEISAGNIYITDN 463

QY 478 THLCFVHTVPWQDLFRNPHQALLHTANRPEDECVGGLACHQLCARGHCWGPQTCVNC 537
DB 464 SNLCYHTTNWTTLFSTINQIRVIRDNKAENCTAEGMVCNHLCSSDGCWGPQDQCLSC 523

QY 538 SQFLRGQECVEECRVLOGLPREVYNARHCLPCHPECOP-ONGSVTCFGEADOCVACAHY 596
DB 524 RFRSRRICIESNLVDGEFREFENGISVCECDPQCEKMDGLTCHGPGPDNCTCKSHF 583

QY 597 KDPFFCVARCPGSKVPDLSPYIWKFPFDEEGACQPCPINCTHSCVDLDDKGC----- 648
DB 584 KQGNPCVCKPCDGLQANSF--IFKYADPDRECHPCPNCTQCNGPTSHDCIYYPTGTH 641

QY 649 ---PACQASPLTSIVSAVV-GILLVVVLGVVFGIILIKRQKIRKYTWRRLLQETELVE 704
DB 642 STLPOQAR-TPL--IAAGVIGGILFVIVGLTFVAVVRRKSIK-KKRALRRFL-ETELVE 696

QY 705 PLTPSGAMPNQAQMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRN 764
DB 697 PLTPSGTAPNQAQRLIKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRN 756

QY 765 TSPKANKEILDEAYVMAGVSPVYRSLGLTCLTSTVQLVTQMLPYPGLCLLDHVRNRRGLG 824
DB 757 TGPKANVEFMDIALIMASMDHPLHVLGVLGCLSPITQLVTQMLPCHCLLEYVHEHKDNG 816

QY 825 SQDLLNWCQIAKGMSELYEDRLVLRHDLAARNVVKSPNHVKITDIFGLARLLDIDETEVH 884

DB 817 SOLLNWCQIAKGMSELYEDRLVLRHDLAARNVVKSPNHVKITDIFGLARLLGDEKEYN 876
QY 885 AUGGKVIKWMALLESILRRFTHQSDVWSVGVVWELMTFGAKPYDGIPIAREIPDLLEKG 944
DB 877 ADGKNPIKWMALDECIIHYRKFTHQSDVWSVGVVWELMTFGKPYDGIPIAREIPDLLEKG 936
QY 945 ERLPOPPICTIDVYIMVVKCMIIDSCRPFRFRELVSFSEFMRMARDPQRFVVIQNEH-LGPA 1003
DB 937 ERLPOPPICTIDVYIMVVKCMIIDADSRPKFELAAEFSEMRMARDPQRYLVIOGDDRMKLP 996
QY 1004 SPLDSTFYRSLLEDMDGDLVDAEYLVFQQGFFCFDDPAPGACGMVHRHRSSTSSGCG 1063
DB 997 SPNDSKFFQNLLEDMDMAEYLVF-QAFNIPPP-----IYTSRARIDSNRS--- 1046
QY 1064 DLTGLGPESEEARPS-----PLAP-SEGAGSDVFDG 1094
DB 1047 --EIGHSPPPATYTPMSGNOFVYRDGFAAEQGVSVYPAPTSTIPEAPVAQGATAE1FDD 1104
QY 1095 DLGMGAAGLQSLPTHDPSPLOQRYSDPTVPLPS-----ETDGYVAPLTCSPQPEYVN 1147
DB 1105 SCNGTILKPVAPHVQEDSTQRYSDPTVFAPERSPRGELDEEGYTPMRDKPKOEYLN 1164
QY 1148 QPDVRRQPPSPREGPLPAARPAAGATLERAKTILSPGKGVYKDVFAFGGAVENPEYITPQG 1207
DB 1165 PVE-----ENPFVSR-----KNGDLQ-----ALDNPYHNASN 1193
QY 1208 GAAQPHPPA-----FSPAFDNLVYWDOP 1233
DB 1194 G-----PKAEDEVYNEPLYNTFANTLGKAEYLNKNTLSMPEKAKKAFDPNPDYWNHSL 1247
QY 1234 PERGA--PPSTFKGTPT-----AENPEYL 1255
DB 1248 PPRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homolog; kinase-related transform
C:Species: Xiphophorus maculatus (southern platyfish)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C:Accession: S06142; S13809
R:Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe
Nature 341, 415-421, 1989
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WIT>
A:Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R:Adam, D.; Maeueller, W.; Scharlt, M.
Oncogene 6, 73-80, 1991
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A:Reference number: S13807; MUID:91125882; PMID:1846957
A:Accession: S13809
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A:Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <WAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 39.0%; Score 2674; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 6.1e-102;

	Matches	575;	Conservative	165;	Mismatches	385;	Indels	152;	Gaps	31;																																																	
Qy	4	AALCRWGLLALLP	GAAST----	QVCTGTDMLRLPAS	PTHETHLDMRLHLRYG	QCVVQGN	59																																																				
Db	8	AALLQ--LLLVL	SIRCCSTDP	DRKVC	CGTSNQMTM---	LDNHYLKM	KWYSGCNVLEN	62																																																			
Qy	60	LELTVLP	NASLSFL	ODIQE	VOGYVLI	IAHQV	RVPQRLR	IVRG	TOLF	FEDNYALAVLDN	119																																																
Db	63	LEITYTO	ENQDLS	FLQIS	IQE	VGYYLI	IAHNV	STIPL	VNLRL	IRQNL	YEGNFTLLVMSN	122																																															
Qy	120	GDPLNNTT	PVTGAS	PGGLRE	LQLRSL	TEIL	AGGVLI	IQRN	POLCY	QD	DTIL	KNDF----	NNF	175																																													
Db	123	YQK--NPSP	----	DVYVG	LKQLQ	LSNL	TEIL	SGV	KVSHN	PLCN	VE	TINM	WDI	VKDT	SNP	179																																											
Qy	176	TVSFWLR	VPKVS	ASAHLEN	R	BRACH	PCSP	CMKGS	RCW	SESS	EDC	QSL	TR	TV	CAGC	-ARCK	234																																										
Db	180	TWNL---	IPHAF	-----	ERQCK	DHGC	VN	GCW	APG	PGCH	CQK	F	KLL	CA	EC	NR	RCR	229																																									
Qy	235	GPLPTDC	CHEO	CAAGCT	CPK	HS	DC	LAC	LHFN	HSGI	CEL	H	CPAL	VTY	NTD	T	FES	MN	PEGR	294																																							
Db	230	GP	KPTDCC	NEHCAG	CT	GP	ATDCL	AC	R	DFND	GD	CT	CP	PP	KI	YI	DV	SH	QV	VDN	PNIK	289																																					
Qy	295	YTFGAS	CVTAC	PYNYLS	T	DVGS	CTV	CP	PLH	NQ	ET	A	DG	T	OR	CE	K	S	K	P	CA	R	V	C	Y	G	L	M	E		354																												
Db	290	YTFGA	ACV	KECP	SNYV	T	E--	GAC	V	R	S	C	A	G	M	L	E	V	D--	ENG	R	S	K	R	C	P	D	G	V	C	P	K	V	C	D	G	I	G	I	347																			
Qy	355	HLRE	VRA	VT	SANI	O	E	F	A	C	K	KI	F	G	S	L	A	P	E	S	F	D	G	P	A	S	N	A	T	A	P	L	Q	P	L	Q	V	F	T	E	L	E	I	T	G		414												
Db	348	SLSNT	IA	VN	ST	NI	R	S	F	S	N	C	T	K	I	N	G	D	I	L	N	R	N	S	F	E	G	D	P	H	Y	K	I	G	T	M	D	P	E	H	L	N	L	I	T	V	K	E	I	T	G		407						
Qy	415	YLYT	S	AW	P	D	S	L	P	D	S	V	F	Q	N	L	O	V	I	R	G	R	I	L	H	N	G	A	S--	L	T	L	Q	L	G	I	S	W	L	G	L	R	S	L	R	E	L	G	S	L	A	L		473					
Db	408	YLV	I	M	N	P	E	N	T	S	L	S	V	F	Q	N	L	E	I	R	G	T	T	F	S	R	G	F	V	V	Q	V	R	H	L	Q	M	L	G	L	R	S	L	R	E	S	A	G	N	V	I		467						
Qy	474	I	H	N	T	L	C	F	V	H	T	P	M	D	O	L	F	R	N	P	H	O	A	L	L	T	A	N	R	P	E	C	V	G	E	G	L	A	C	H	O	L	C	A	R	G	H	C	W	G	P	T	Q		533				
Db	468	L	K	N	T	L	Q	L	R	A	N	T	I	N	W	R	L	F	R	S	E	D	S	I	E	D	A	T-----	E	N	Q	T	C	N	N	E	S	E	D	G	C	W--	P	G	P	T	M		519										
Qy	534	C	V	N	C	S	Q	F	L	R	G	O	E	V	E	E	C	R	V	L	O	G	L	P	R	E	Y	N	A	R	H	C	L	P	C	H	P	E	Q	P	O	N	G	S	V	T	C	F	G	E	A	D	O	C	V	A	C		593
Db	520	C	V	S	C	L	H	V	D	R	G	R	C	V	A	S	C	N	L	L	O	G	E	P	R	E	A	Q	V	D	G	R	C	V	O	C	H	E	C	L	V	O	T	D	S	L	T	C	V	G	P	A	N	C	S	K	S		579
Qy	594	A																																																									

Qy	1074	EEAPRSLPASEGAGSDVDFDGLGMGAAGLQSLPHDRSPLOQRYSEDPTV-PLPSETDG	1132
Dd		: : : : :	
Dd	1020	-----EPCIDPTGH-----PVRENSITLRNISDPONALEDKLDG	1054
Qy	1133	VYAPLTCSPOPEYVNQDVRQPQ-----PSPRE-----GPLP-AARPAGATLERAKTL	1179
Dd		: : : : : : :	
Dd	1055	H-----EYNQPGSTSRLSDIYNPNTEYDLTGWGPVSLSSQEAEINFSRPEYL	1104
Qy	1180	SPKGNGVVVDVFAGFAGVENPEYLTPOGGAAQPQHPPPAFSPAFDNLYYWDQPPERCAP	1239
Dd		: : : : : : : : : : :	
Dd	1105	NTNQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGAL	1139
Qy	1240	PSTFKGPTTAENPEYLG	1256
Dd		: : :	
Dd	1140	TGNMGFLPAAALEYLG	1156

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Isling, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the EBBB/epidermal growth factor receptor gene family
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whittney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, P.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor
A:Reference number: I59164; MUID:90311312; PMID:2164210
A:Accession: I59164
A>Status: preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:I19880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match	35.3%	Score	2418.5	DB 2	Length	1342			
Best Local Similarity	40.4%	Pred. No.	1.8e-91						
Matches	532	Conservative	192	Mismatches	458	Indels	135	Gaps	32
QY	10	GLLLALLPPCAA--STQVCTGTDWKRLRLPASPTHLDMLRHLYGCGVCOVGNLELTYLPT	67						
DB	11	GLLFLSARGSEVGNSSQAVCPETLNGLSVTGDAENQVOTLYKLFEREVVNGNLEIVLVTGH	70						
QY	68	NASLSFLQDIQEOGYVYLIAHNRQVQVPLQRLRLRIVRGQTQLFEDNYALAVLDNGDPLNNTT	127						
DB	71	NADLSFLQWIREVTGYVLVANNEFSTPLPLPNLRAVRGTQVYDYGKFAIVM-----LNYNT	125						
QY	128	PVTGASPGGLRELQRLSRSLTEILKGGVLIQBNPOLCQYDITLWKDPNNFTVSFWLRVPKVS	187						
DB	126	-----NSSHALRQLRLTQLTEILSGGYVIEKKDKLCHMDTDMRDIVR-----DRDA	172						
QY	188	ASHLNRSRACHPCSPMKSGSRGWSSESDCQLSTRVCAGGC-ARCKGPLPTDCCHEQC	246						
DB	173	EIVVKNGRSCPPCHEVCKG-RCWFGSESDCQLTKTICAPQCNCHGCFGPNPQCCHDEC	231						
QY	247	AAGCTGPKHSDDCLACHFNHSHGICELHCPALVYNTDITFESMPNPBGRYTFFGASCVTACP	306						

Db 578 HFRDGPCHVMSCPHGILG--AKGPIYKYPDAQNECRPCHEMTCQCNCPGELQDCIGQAEV 635
Qy 653 RASPLTISVAVGILLVVLGVVFGIILKRQOKIR-KYTMRRLLQETELVEPLTPSGA 711
Db 636 LMSRPHLVIATVG--LAVILMLGSGFLYWRGRIOKRAMRRLYERGESIEPLDPS-E 692
Qy 712 MPNQOMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
Db 693 KANKVLARIFKETELRKVLKLGSGVFTVHKGIWIPGEGSIKIPVCIKVIKDSQGRSQFO 752
Qy 772 EILDEAVYVAGVSGPYVSRLLGICLTSTVOLVTLMPYGCCLLDHVRNRRGLSGQDLNW 831
Db 753 AVTDHMLAVGSDHAHIVRLGLCPGSSQLVQYPLGSLLDHVKQHRETLPQQLLNW 812
Qy 832 CMQIAKMSYLEDVRLHVRDLAARNVLPKSNHVKITDFGLARLLDIDETEHADGKGVP 891
Db 813 GVQIAKGYMYLEHSMVHRDLARNVLMKSPSQVQVADGVADLLPPDDKQLLHSEAKTP 872
Qy 892 IKWMALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPPOP 951
Db 873 IKWMALESIHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLAELIPDLEKGERLAOPQ 932
Qy 952 ICTIDVYMWKCMWIDSECRPRPRELVSEFSRMDRDPORFVVIQNEEDLGPASPLDSTFY 1011
Db 933 ICTIDVYMWKCMWIDENIRPTFKELANEFTRMARDPRLYLVIKRAS-GPGTP--PAAE 989
Qy 1012 RSLDEDDMGDLVDAEYLVPOQGFPCPDPAFGAGGVHRRHRSSTRSGGDLTLGLEP 1071
Db 990 PSVLTTEL-----QEAELEPEL-----DLDLLEA 1015
Qy 1072 SEE-----EAPRSPAPSEG-----AGSDVFDGLGMAAGK 1103
Db 1016 EEEGLATSLGALSALPTGTLTRPGQSLSPSSGYMPMNOSSLGEACLDASVAGRGREF 1075
Qy 1104 LQSLPHTDPSPLQRYSEDPTVPLPSETGYV---APL-----TC-----SPOPE--- 1144
Db 1076 SRPISLH-PIPRGR-----PASESEGHVTGSEAELOEKVSVCRSRSRSPRGRDS 1127
Qy 1145 -YVNPQDVRPOPSPREGP-----LPAARPAGATLERAKTLP-SRNGVYV----- 1187
Db 1128 AYHSQRHSLLTPTVPLSPGLEEDGNGVMPDTHLRGASSSREGTLSSVGLSSVLGTTEE 1187
Qy 1188 --KOVFAFGGAVENPEYLTPOGGAAPQPHPP 1216
Db 1188 EDED-----EEYEYMNKRGRGSP-PRPP 1209

RESULT 11

TVFVLV
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and product formation in the presence of the src gene
A:Reference number: A00643; PMID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
A:Cross-references: GB:M10066; GB:M13801; NID:g211749; PIDN:AAA48763.1; PID:g211750
A:Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific protein kinase
F:1-6/Product: gag protein (fragment) #status predicted <GAG>
F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase envB #status predicted <ERB>
F:194-459/Domain: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred No. 3.4e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
Qy 584 GPEADQCVACAHYKDPFPCVACRPSGVKPDLSYMPIMKFPDEBEGACQPCPINTCHSCVDL 643
Db 60 GP--DHCMKAFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNCNTRCKGK 116
Qy 644 DDGCPAPQASPLTSIVSAVV-GILLVVLGVVFGIILKRQOKIRKVTMRLLQETEL 702
Db 117 GLEGCP--NGSKTPTSIAGVVGGLCLVWGLGIGLYLRRR-HIVRKTRRLRLOEREL 172
Qy 703 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 762
Db 173 VEPLTPSGEAPNQAHRLILKETEFKVKVVLGSGAGFTVYKGIWIPDGENVKIPVAIKELR 232
Qy 763 ENTPSKANKEILDEAYVAGVSPVSRLLGICLTSTVOLVTLMPYGCCLLDHVRNRR 822
Db 233 EATSPKANKEILDEAYVNASVDNPHVCRLLGICLTSTVOLITOLMPYGCCLLDVIREHKDN 292
Qy 823 LGSODLLNMCQIAKMSYLEDVRLVHRDLAARNVLPKSNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLNMCQIAKGMNLEERLVRDLAARNVLPKTPQHVKITDFGLAKLLGDAEKE 352
Qy 883 YHADGKVPYIKWMALESILRRRFTHQSVDWSYGVTVWELMTFGAKPYDGIPIAREIPDLE 942
Db 353 YHAEKGKVPYIKWMALESILHRIYTHQSVDWSYGVTVWELMTFGSKPYDGIPIASEISSVLE 412
Qy 943 KGERLPQPPICITIDVYMWKCMWIDSECRPRPRELVSEFSRMDRDPORFVVIQ-NEDLG 1001
Db 413 KGERLPQPPICITIDVYMWKCMWIDADSRPKFRELIAEFSKWARDPPRYLVIVLQGERMH 472
Qy 1002 PASPLDSTFYSLLEDDDMGDLVDAEYLVPOQGFPCPDPAFGAGGVHRRHRSSTRSG 1061
Db 473 LPSPTDSKPYRTLMEEDEMDIVDADEVLPVHQGF-----NSPST--- 513
Qy 1062 GGDLTGLGLEPSEEAAPRSP-----APSEGAGSDVFDGDLGMAAGKLSLPHTHDPSPLQ 1116
Db 514 -----SRTPLLSSLSATSNNSATNCID-----RNGQHPVRESFVQ 550
Qy 1117 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKPS----- 585
Qy 1175 RAKTLPQNGVGVKQVDF-----AFGAVENPEYLTPOGGAAPQPHPPAP 1219
Db 586 ----TAMVQNCIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1220 SPAPDNLYYWDQ-----DPPE-----RGAPPSTFGTPTAENPEYVLGLDVP 1260
Db 634 KTVFESSPYWIOQSGNHQINLNDPDIQQDFLPNETKPNGLLKVPAAENPEYLRVAAP 689

RESULT 12

TVYUWH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A:Reference number: A00644; PMID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
R:Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Marti
Science 224, 1456-1459, 1984
A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type
A:Reference number: A38022; PMID:84223957; PMID:6328658

C:Superfamily: epidermal growth factor receptor; protein kinase homology										
C:Keywords: ATP; phosphotransferase										
F:135-400/Domain: protein kinase homology <KIN>										
F:143-151/Region: protein kinase ATP-binding motif										
Query Match		23.9%	Score	1640;	DB 2;	Length	545;			
Best Local Similarity		54.9%	Pred. No.	3.8e-60;						
Matches		345;	Conservative	69;	Mismatches	122;	Indels	92;	Gaps	15;
Qy	584	GPEADQCACAHYKDPFCVAPRCVSGVKPDLSYMPKFPDEGACQPCPINCTHSCVDL	643							
Db	1	GP--DHCWKCAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP	57							
Qy	644	DDKGCPEAQRASPLTSIVSAV-GILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETEL	702							
Db	58	GLEGCP--NGSKTPSIAAGVVGCLCLVVGVLGILGLYLR--HIVKRTLRLRLQEREL	113							
Qy	703	VEPLTSGAMPNQAQRIILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPIVAIKVL	762							
Db	114	VEPLTSGEAPNQAHLRIILKETEFKKVKVLGFGAFGVYKGLWIPGEKVTIPIVAIKEL	173							
Qy	763	ENTSPKANKEIIDEAYVMAGVSGPYVSRLLIGICLTSTVOLTMPYGCLLDHYVNRGR	822							
Db	174	EATSPKANKEIIDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPYGCLLDYIREHKON	233							
Qy	823	LGSQDLLNWCQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE	882							
Db	234	IGSQYLLNWCQIAKGMNLEERHLVHRDLAARNVLVKTPQDVKITDFGLAKQLGADKE	293							
Qy	883	YHADGKVPIKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLE	942							
Db	294	YHAEGKVPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLE	353							
Qy	943	KGERLPQPPICTIDVTMIMVKMIDSECRPRRELVSFSPRMARDPQRFVVIQ-NEDLG	1001							
Db	354	KGERLPQPPICTIDVTMIMVKMWDADSPKPRELIAEFSKWARDPPRYLVLIQDERMH	413							
Qy	1002	PASPLDSTFYRSLLDDMGDLVDABEYLVPOQGFPCPDPAAGAGMVRHRRSSSTRSG	1061							
Db	414	LPSPTDSKFYRTLMEEDMEDIVDADAYLVPHQGF-----NSPST-----	454							
Qy	1062	GGDLTLGLRPSSEEA PRSPL-----APSEGAGSDVFDGLGMCAAKGLQSLPTHPSPLQ	1116							
Db	455	-----SRTPLLSLSATSNNSATNCIDRNGG-----H-----	481							
Qy	1117	RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER	1175							
Db	482	-----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNIYNYISLTAISK	523							
Qy	1176	AKTLPCKNGVVKDVFAGGAVENPEYL	1203							
Db	524	LPMDSRVQN-----SHSTAVDNPEYL	544							

Search completed: July 22, 2003, 09:08:44
Job time : 32.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec
Title: SEQ4-149-163-14
Perfect score: 6848
Sequence: 1 MELAALCRWGLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description	
Result No.	Score	Query Match	Length	DB ID		
1	6709	98.0	1255	1	ERB2 HUMAN	P04626 homo sapien
2	5924	86.5	1257	1	ERB2 RAT	P06494 rattus norv
3	5914.5	86.4	1254	1	ERB2 MESAU	P00553 mesocricetu
4	3162	46.2	1210	1	EGFR HUMAN	P00533 homo sapien
5	3142	45.9	1210	1	EGFR MOUSE	Q01279 mus musculu
6	2972.5	43.4	1308	1	ERB4 HUMAN	Q15303 homo sapien
7	2959	43.2	1308	1	ERB4 RAT	Q62956 rattus norv
8	2690.5	39.3	1167	1	MXRK XIPMA	P13388 xiphophorus
9	2427.5	35.4	1342	1	ERB3 HUMAN	P21860 homo sapien
10	2364.5	34.5	1339	1	ERB3 RAT	Q62799 rattus norv
11	1964	28.7	1426	1	EGFR DROME	P04412 drosophila
12	1749.5	25.5	634	1	ERBB ALV	P00534 avian leuko
13	1703	24.9	604	1	ERBB AVIER	P00535 avian eryth
14	1630	23.8	540	1	ERBB AVIEU	P11273 avian eryth
15	1611	23.5	703	1	EGFR CHICK	P13387 gallus gall
16	1302	19.0	1323	1	LT23 CAEEL	P24348 caenorhabdi
17	1142.5	16.7	245	1	ERB2 MOUSE	P70424 mus musculu
18	734	10.7	1363	1	ILPR BRALA	Q02466 branchiosto
19	714	10.4	1297	1	IRR HUMAN	P14616 homo sapien
20	714	10.4	1300	1	IRR MOUSE	Q9wt14 mus musculu
21	707	10.3	1382	1	INSR HUMAN	P06213 homo sapien
22	699.5	10.2	1300	1	IRR CAVPO	P14617 cavia porce
23	698	10.2	1383	1	INSR RAT	P15127 rattus norv
24	697	10.2	1607	1	IMPR LYNST	Q25410 lymnaea sta
25	696.5	10.2	1372	1	INSR MOUSE	P15208 mus musculu
26	665	9.7	1477	1	HTK7 HYDAT	Q25197 hydra atten
27	649	9.5	1367	1	IG1R HUMAN	P08069 homo sapien
28	640	9.3	1373	1	IG1R MOUSE	Q60751 mus musculu
29	636.5	9.3	1370	1	IG1R RAT	P24082 rattus norv
30	622	9.1	1390	1	INSR AEDAE	Q93105 aedes aegypt
31	607	8.9	987	1	EPB4 HUMAN	P54760 homo sapien
32	591.5	8.6	984	1	EPB1 CHICK	Q07494 gallus gall
33	591	8.6	2146	1	INSR DROME	P09208 drosophila

ALIGNMENTS									
RESULT 1									
ERB2 HUMAN									
ID	ERB2_HUMAN	STANDARD;	PRT;	1255	AA.				
AC	P04626;								
DT	13-AUG-1987	(Rel. 05, Created)							
DT	13-AUG-1987	(Rel. 05, Last sequence update)							
DT	15-JUN-2002	(Rel. 41, Last annotation update)							
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)								
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell								
DE	surface receptor HER2) (MLN 19).								
GN	ERBB2 OR HER2 OR NGL OR NEU.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
OX	NCBI_TaxID=9606;								
RN	[1]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=86118663; PubMed=3003577;								
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,								
RA	Saito T., Toyoshima K.;								
RT	"Similarity of protein encoded by the human c-erbB-2 gene to								
RT	epidermal growth factor receptor.";								
RL	Nature 319:230-234(1986).								
RN	[2]								
RP	SEQUENCE FROM N.A.								
RX	MEDLINE=86070181; PubMed=2999974;								
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,								
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,								
RA	Frankie U., Levinson A., Ullrich A.;								
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor								
RT	shares chromosomal location with neu oncogene.";								
RL	Science 230:1132-1139(1985).								
RN	[3]								
RP	SEQUENCE OF 737-1031 FROM N.A.								
RX	MEDLINE=86016729; PubMed=2995967;								
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;								
RA	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the								
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a								
RT	human salivary gland adenocarcinoma.";								
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).								
RN	[4]								
RP	VARIANTS VAL-654 AND VAL-655.								
RX	MEDLINE=93194196; PubMed=8095488;								
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;								
RT	"Characterization of a new allele of the human ERBB2 gene by allele-								
RT	specific competition hybridization.";								
RL	Genomics 15:426-429(1993).								
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,								
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A								
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-								
CC	ALPHA AND AMPHIREGULIN.								
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein								
CC	tyrosine phosphate.								
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS								
CC	(POTENTIAL).								
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.								

34	589.5	8.6	977	1	EPB2_MOUSE	Q03145 mus musculu
35	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
36	585	8.5	987	1	EPB4_MOUSE	P54761 mus musculu
37	584.5	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
38	577.5	8.4	985	1	EPBA_XENLA	Q91571 xenopus lae
39	574.5	8.4	984	1	EPB1_RAT	P09759 rattus norv
40	573.5	8.4	1053	1	PAK1_CHICK	Q00944 gallus gall
41	570	8.3	902	1	EPBB_XENLA	Q91736 xenopus lae
42	569	8.3	1068	1	PAK1_XENLA	Q91738 xenopus lae
43	568.5	8.3	984	1	EPB1_HUMAN	P54762 homo sapien
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.2	1052	1	PAK1_MOUSE	P34152 mus musculu


```

QY 717 QMRILKTELKVKVGLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKKILDE 776
DB 711 QMRILKTELKVKVGLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKKILDE 770
QY 777 AYVMAGVGSYPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSODLLNWCQIA 836
DB 771 AYVMAGVGSYPVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSODLLNWCQIA 830
QY 837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMA 896
DB 831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYADGGKVPKMA 890
QY 897 LESILRRFRHQSDVNSGYTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 956
DB 891 LESILRRFRHQSDVNSGYTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIMVKCMIIDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
DB 951 VYMIMVKCMIIDSECRPRFRELVSERFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
QY 1017 DDMGDLVDAEYLVPOQGFCDPAPAGCMVHRRSRSSSTRSGGDLTLGLEPSEEA 1076
DB 1011 DDMGDLVDAEYLVPOQGFCDPAPAGCMVHRRSRSSSTRSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGDLGMAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGDLGMAKGLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFVAFGGA 1196
DB 1131 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVDFVAFGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPQHPHPPAPSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 2
ERB2_RAT
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC F06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
```

```

[3]
RN STRUCTURE BY NMR OF 650-668.
RP MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X03362; CAA27059.1; ALT_INIT.
CC PIR: A24562; TVRTNU.
CC HSPSP; P11362; IFGK.
CC InterPro: IPR000494; EGFR_L domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC InterPro: IPR004019; YLP_motif.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L domain; 2.
CC Pfam: PF02757; YLP_2_ domain; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
CC SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```

```
FT DISULFID 517 530 BY SIMILARITY.
FT FT 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT FT 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT FT 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 W -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.5%; Score 5924; DB 1; Length 1257;
Best Local Similarity 86.7%; Pred. No. 3.3e-308;
Matches 1097; Conservative 51; Mismatches 104; Indels 14; Gaps 5;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGIAGTQCTGCTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLIHANQVQVPLQRLRVGTQQLFEDNYALAVLDNG 120
Db 61 ELTYVPANASLSFLQDIQEVGYVLIHANQVQVPLQRLRVGTQQLFEDNYALAVLDNR 120
Qy 121 DPLNNTTTPVT-GASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFT 176
Db 121 DPQDNVAASPTGRTPEGLRELQLRSLTEILKGGVLIIRGNPOLCYQDNVWLKDVFRKN-- 178
Qy 177 VSWLRVPKVSASHLENRSRACHPCSPMKSGRSRCHWGSSEDCQSLTRTVTCAGGCARCKGP 236
Db 179 -----QLAPVDID--TNRSRACPPCAPACKDNHGWGESPEDCQILTGITCTSGCARCKGR 231
Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYT 296
Db 232 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYT 291
Qy 297 FGASCVTACPNYLSLTVSGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMEHL 356
Db 292 FGASCVTTCPNYLSLTVSGSCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLGMEHL 351
Qy 357 REVRAVTSANIQEPAGCKKIPGSLAFLPESPDGDPASNTAPLQEQLVQVFTLEETGYL 416
Db 352 RGAARITSDNVQVEPDGCKKIFGSLAFLPESPDGDPSSGIAPLRPEQLQVFTLEETGYL 411
Qy 417 YISAWPDLPLSVFQNLQVTRGRILHNGAYSLTLQGLISWGLSLRELGSGLALIH 476
Db 412 YISAWPDLRLSVFQNLRIIRGRILHHDGAYSLTLQGLISWGLSLRELGSGLALTHR 471
Qy 477 NTHLCFVHTVPWDLFRNPHQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPGTQCV 535
Db 472 NAHLCFVHTVPWDLFRNPHQALLHSGNRPEDELCVSSGLVCSLCAHGCWGPGTQCV 531
Qy 536 NCSOFLRGQECVECRVLQGLPREYNARHCLCPHCEQPONGSVTCFGEADQVCAH 595
Db 532 NCSHFLRGQECVECRVWKGLPREYVSDKRLCPHCEQPONSSETCFGEADQVCAH 591
Qy 596 YKDPFPCVAPCSPKPDLSYMPYLPKPDDEGACQPCINCTHSCVDLDDKGCBAORAS 655
Db 592 YKDSSSCVAPCSPKPDLSYMPYLPKPDDEGICQPCINCTHSCVDLDERGCPAORAS 651
Qy 656 PLTSIVSAVGLILVVLGVVFGILIKRRQKIRKYMRLLOETELVELPLTPSGAMPNQ 715
Db 652 PVTFIATVVGVLFLILVVGILIKRRQKIRKYMRLLOETELVELPLTPSGAMPNQ 711
Qy 716 AQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKBILD 775
```

```
Db 712 AQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKBILD 771
Qy 776 EAYVMAGVSPYVSRLLIGICLTSTVQLVTQMPYGCCLLDHVNRGRGLGSDLLNWCQI 835
Db 772 EAYVMAGVSPYVSRLLIGICLTSTVQLVTQMPYGCCLLDHVNRGRGLGSDLLNWCQI 831
Qy 836 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKGVPIKWM 895
Db 832 AKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKGVPIKWM 891
Qy 896 ALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOICTI 955
Db 892 ALESILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOICTI 951
Qy 956 DVYIMVYKCMIDSECPRELVSFESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1015
Db 952 DVYIMVYKCMIDSECPRELVSFESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1011
Qy 1016 EDDMDGLVDAAEYLVPOQGFCCPDPAFGAGGMVHRRSSSTRSGGDLTLGLEPSEEE 1075
Db 1012 EDDMDGLVDAAEYLVPOQGFCCPDPAFGAGGMVHRRSSSTRSGGDLTLGLEPSEEG 1071
Qy 1076 APRSPLAPSEAGSDVFDGDLGMGAAGKLSLPTHPSPQLQRYSEDPVPLPSETDGYVA 1135
Db 1072 PPRSPLAPSEAGSDVFDGDLGMGAAGKLSLPTHPSPQLQRYSEDPVPLPSETDGYVA 1131
Qy 1136 PLTCSPOPEYVQDPVPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFG 1195
Db 1132 PLACSPQPEYVQSEVQPPPLTPEGLPVPVRPAAGATLERPKTILSPGKNGVYKDVFAFG 1191
Qy 1196 AVENPEYLTTPQGAAPQHPHPPAFSPAFDNLVYWDQDPPERPAPPTFKGTPTAENPEYL 1255
Db 1192 AVENPEYLVREGTASPSPHSPAFSPAFDNLVYWDQDPPERPAPPTFKGTPTAENPEYL 1251
Qy 1256 GLDVPV 1261
Db 1252 GLDVPV 1257

RESULT 3
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-253(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
```

CC RESIDUES.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D16295; BAA03801.1; -

DR HSSP; P11362; LFCK.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_pkinase.

DR InterPro; IPR004019; YLP motif.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF02757; YLP; 2.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; Tyrc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.

DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;

KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.

FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 653 675 POTENTIAL.

FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 158 168 CVS-RICH.

FT DOMAIN 472 644 CVS-RICH.

FT DOMAIN 720 987 PROTEIN KINASE.

FT NP_BIND 726 734 ATP (BY SIMILARITY).

FT BINDING 753 753 ATP (BY SIMILARITY).

FT ACT_SITE 845 845 BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.

FT DISULFID 199 212 BY SIMILARITY.

FT DISULFID 236 244 BY SIMILARITY.

FT DISULFID 240 252 BY SIMILARITY.

FT DISULFID 255 264 BY SIMILARITY.

FT DISULFID 268 295 BY SIMILARITY.

FT DISULFID 299 311 BY SIMILARITY.

FT DISULFID 315 331 BY SIMILARITY.

FT DISULFID 334 338 BY SIMILARITY.

FT DISULFID 511 520 BY SIMILARITY.

FT DISULFID 515 528 BY SIMILARITY.

FT DISULFID 531 540 BY SIMILARITY.

FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.

FT DISULFID 567 584 BY SIMILARITY.

FT DISULFID 587 596 BY SIMILARITY.

FT DISULFID 600 623 BY SIMILARITY.

FT DISULFID 626 634 BY SIMILARITY.

FT DISULFID 630 642 BY SIMILARITY.

FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 658 658 V -> E (IN ONCOGENIC NEU).

FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).

FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 86.4%; Score 5914.5; DB 1; Length 1254;
Best Local Similarity 86.4%; Pred. No. 1.1e-307;
Matches 1092; Conservative 59; Mismatches 100; Indels 13; Gaps 4;

QY 1 MELAALCRWGLLLALLPQCAASTOVCTGDMKLRLPASPETHLDMLRLHYQGVQVQGNL 60
DB 1 MELAAWCGMLLLALLSPGASGTQVCTGDMKLRLPASPETHLDIVRHLYQGVQVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQCVYLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPANATLSFLQDIOEVQCVYMLIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNR 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDQTLWKDF---NNFTV 177
DB 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIQORNPOLCYQDQTLWKDFVRKNN-- 177
QY 178 SFWLAVPKVSAHLENRSRACHPCSPMKGSCWGESSEDCQSLRTVCAGGCARCKGPL 237
DB 178 ----QLAPVDID--TNRSRACPPCAPACKDNHCWASPEDCQTLTGTIAPRAVPAARAL 231
QY 238 PTDCCHCOCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTF 297
DB 238 PTDCCHCOCAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTF 291
QY 298 GASCVTACPNYNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLR 357
DB 292 GASCVTTCPNYNYLSTEVGSCTLVCPLNHQEVTAEDGTQRCCKSKSCARVCYGLGMEHLR 351
QY 358 EVRAVTSANIOEFAGCKKIFGSLAFLPSPGDDPASNTAPLOPEQLOVETLEEITGYLY 417
DB 352 GARAITSANIOEFAGCKKIFGSLAFLPSPGDDPASNTAPLOPEQLOVETLEEITGYLY 411
QY 418 ISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIHNN 477
DB 412 ISAWPDSLHDLVSFQNLQVIRGRIILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIHNN 471
QY 478 THLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNC 537
DB 472 THLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNC 531
QY 538 SFLRGQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYK 597
DB 532 SHFLRGQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCVACAHYK 591
QY 598 DPPFCVAPCPGKPDLSYMPIWKPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPL 657
DB 592 DSPFCVAPCPGKPDLSYMPIWKPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPA 651
QY 658 TSIVSAVVGILLVAVGVFGLIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAQ 717
DB 652 TSIIATVVGILLVAVGVFGLIKRRQKIRKYMRLLOETELVEPLTPSGAMPNQAQ 711
QY 718 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEA 777
DB 712 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDEA 771
QY 778 YVMAGVSPYVSRLLIGICLTSTVQLVLTQMPYVGLLDHVRNRRGLSGDGLNWCQIAK 837
DB 772 YVMAGVSPYVSRLLIGICLTSTVQLVLTQMPYVGLLDHVRNRRGLSGDGLNWCQIAK 831
QY 838 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 897
DB 832 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 891
QY 898 ESILARRRTHOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 957
DB 892 ESILARRRTHOSDWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 951
QY 958 YIMVVKCWMIDSECRPRELVSFSESRMAROPQRFVVIQNEIDLPASPLDSTFYFSLLED 1017
DB 952 YIMVVKCWMIDSECRPRELVSFSESRMAROPQRFVVIQNEIDLPASPLDSTFYFSLLED 1011

OY 1018 DMGDLVDAEYLVPOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEP 1077
 DB 1012 DMGDLVDAEYLVPOGFFFPDPAPGAGSTAHRRSSSTRSGGDLTLGLEPSEEP 1071
 OY 1078 RSLAPSEGAGSDVFDGLGMAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGVYAPL 1137
 DB 1072 RSLAPSEGAGSDVFEGLGMAATKQPSIPRDLSPQLQRYSEDPTVPLPSETDGVYAPL 1131
 OY 1138 TCSPOEYVNOVDVPOPPSPREGPLPAARPAAGATLERAKTISPGKNGVVKDVFAGGAV 1197
 DB 1132 ACSPOEYVNOQVPRPQPLPETEGPLPPVRPAGATLERKTLSPGKNGVVKDVFTEGGAV 1191
 OY 1198 ENPEYLTPOGGAAPQPHPPAPFSPAFDNLNLYWDQDPPERCAPPSTFKGTPTAENPEYLG 1257
 DB 1192 ENPEYLVPRGSSASQPH-PPALCPAFDNLNLYWDQDPSERGPSNPTEGPTAENPEYLG 1250
 OY 1258 DVPV 1261
 DB 1251 DVPV 1254
 RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 ID EGFR_HUMAN AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Maihle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Garti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RC TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Maihle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,
 RA Maihle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor


```

QY 959 MIMVKCMIDSECRPRELSEFSEFMRDQPFVVIQ-NEDLGPASPLDSTFYRSLLED 1017
DB 945 MIMVKCMIDADSRKFKELIIEFSKMRDQPVLYIQGDERMHLPSPTDSFYRALMDE 1004
QY 1018 DMGDLVDAEYLYVQOQFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAR 1077
DB 1005 EDMDDVDVADEYLYVQOQFF-----SSPSTS 1030
QY 1078 RSLAPSEAGSDVFDGLGMAAGKLOSLPHDPSPLQRYSEDTPVLPSET--DGVA 1135
DB 1031 RTPLLSSLSATSN--NSTVACIDRNLQSCPKEDSFLORYSDTGTALTEDSIDTFL- 1087
QY 1136 PUTCSPOEYVNPQVRPOPPSPREGPLPAARPAGATLERAKTLSPGKNGVYVFAFGG 1195
DB 1088 -----PVPEYINQ-SVPKRPAGSVQNPVHNPQNP-----APSRDPHYQD--PHST 1131
QY 1196 AVENEYLY-TPOGGAAPQHPHPPAFSPADNLYYNDQ-----DP-----PERGA 1238
DB 1132 AVGNEYYLNTVO-----PTCVNSTFDSPAHWAQKGSHOISLDNPDYQDFFPKPEAK 1182
QY 1239 PPSTFKGTPTAENPEYL 1255
DB 1183 PNGIEFGS-TAENAEYL 1198

RESULT 5
ID_EGFR MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avioli A., Lux I., Ullrich A., Schlessinger J., Givol D., Morse B.;

```

"Comparison of EGF receptor sequences as a guide to study the ligand binding site";
 Oncogene 6:673-676(1991).
 [6]
 SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Eisinger D.P., Serrero G.;
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
 AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC SUBCELLULAR LOCATION: Type I membrane protein.
 CC MISCELLANEOUS: Binding of EGF to the EGF-receptor leads to
 induction, internalization of the EGF-receptor complex,
 induction of the tyrosine kinase activity, stimulation of cell DNA
 synthesis, and cell proliferation.
 CC SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 or send an email to license@sib-sib.ch).

 CC EMBL; X78987; CAA55587.1; -
 DR EMBL; U03425; AAA17899.1; -
 DR EMBL; X59698; CAA42219.1; -
 DR EMBL; L06864; AAA53029.1; -
 DR EMBL; Z12608; CAA78249.1; -
 DR HSSP; P11362; 1FGK
 DR MGI; MGI:95294; Egfr.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FY; 3.
 DR SMART; SM00219; TyrK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT DOMAIN 714 981
 FT NP_BIND 720 728
 FT BINDING 747 747
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 326
 FT DISULFID 329 333

FT	NP_BIND	724	732	ATP (BY SIMILARITY).
FT	BT_BINDING	751	751	ATP (BY SIMILARITY).
FT	ACT_SITE	843		BY SIMILARITY.
FT	DISULFID	189	197	BY SIMILARITY.
FT	DISULFID	193	205	BY SIMILARITY.
FT	DISULFID	213	221	BY SIMILARITY.
FT	DISULFID	217	229	BY SIMILARITY.
FT	DISULFID	230	238	BY SIMILARITY.
FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPIC	626	648	NGPSTSDHCYIPWGHSTLPQHA -> IGSSEDICGLMD
FT				(IN ISOFORM JM-BI).
SEQ	SEQUENCE	1308 AA; 146807 MW; 554AE80985D88761 CRC64;		
Query Match				
Best Local Similarity 43.4%; Score 2972.5; DB 1; Length 1308;				
Matches 613; Conservative 185; Mismatches 369; Indels 191; Gaps 31;				
Qy	9	WGLLLLLPPGAA-----STQCTGTDMKRLPASPETHLDMLRHLGYGCGVQGNLBTY	64	
Db	8	WVWVSLTVAAGTVQPSDSQSVCAGTENKLSLSLDLEQQYRALRKYVENCEVWVGNLEITS	67	
Qy	65	LPTNASISFLQDIQEQGVYVLIAHNOVROVPLQRLRIVRGQTQGFEDNYALAVLDNGDPLN	124	
Db	68	IEHNRDLFLRISVREVTGVYVVALNFRYLPLENRLIRGTUKLYEDYALAFILNYRKDG	127	
Qy	125	NTTPTVTGASPGGURELQLRSLTEILRGGVLIQRNPOLCYODTILWKDF-----NNFTVS	178	
Db	128	NF-----GLQELGLKNLTEILNGVGVVDQNKFLCYADTIHWQDIVRNWPNSLTL-	177	
Qy	179	FWLRVPKVSASHLENSRACHPCSPCKSGRCHGESSEDCQSILTRTVCAGC-ARCKGPL	237	
Db	178	-----VST-----NGSSGCGRCHKSGCTG-RCWGPTEHNCQTLTRTVCAQCDCGRCYGPY	225	
Qy	238	PTDCHEQCAAGCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTFESMPNPEGRYTF	297	
Db	226	VSDCHRECAAGCGSPKDTDCFAFMFNUSGACVTCQPTQTFVTNPTTFLEHFNNAKYT	285	
Qy	298	GASCVTACPVNYLSTDVGSCCTLVCLPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLR	357	
Db	286	GAFCVKCKPHNFV-VDSSSSCVRACPSCKMEV-EENGIKMKCPCTDICPKACDGGTGSML	343	
Qy	358	EVRAVTSANIQEAFGCKKIFGSLAFILPESFDGDPASNTAPLQEQIQLVETLEITCYLY	417	

```
Db 344 SAQTVDSSNIDKFNCTKINGNLIFLVTGIHGDYPNAIEAIDPEKLVNFTVREITGFLN 403
Qy 418 ISAWPDSLPDLVSFQNLQVIRGRIHLHNGAYSITLQGLGSIWGLRSLRELGSGLALIHNN 477
Db 404 IQSWPNNTDFSVFNLVTIGRVLVYGLSLILKQOGITSLQFQSLKEISAGNIYITDN 463
Qy 478 THLCFVHTVPDOLFRNPHOALLHTANRDECEVGEGLACHOLCARHGCHGPGPTQCVNC 537
Db 464 SNLCYHTINWTLFTSTINQIRIVIRDNKAENCAETAGWCVNHLCSGDCGCGPGPOQCUSC 523
Qy 538 SOFLRQGECEVCEVRLQGLPREYVNAHCLPCHPECQP-QNGSVTCFGEADQCACAHY 596
Db 524 RFRSGRICIESCNLYDEGEFENGISICVECDPQCEKXMEDGLLTCHGPGPNDCTKSHF 583
Qy 597 KDPPFCVACRSGVAPDLSYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGC----- 648
Db 584 KDGPNCEKCPDGLQANSF--IFKYADPDRECHPCPNCTQCGNGPTSHDCIYYPWTGH 641
Qy 649 ---PAEQRASPLTISVSAW-GILLVWVGVVFGILIKRQOKIRKYTMRRLLQTELVE 704
Db 642 STLPOHAR-TPL--IAAGVIGGLFVLIVGLTFVAVYVRKSIK-KKRALURRL-ETELVE 696
Qy 705 PLTPSGAMPNOQMRLKETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLREN 764
Db 697 PLTPSGTAPNOAQLRKETELKRVKVLGSGAGFTVYKGIWPEGETVKIPVAIKILNET 756
Qy 765 TSPKANKEILDYAWMAGVSGPYVSRLLIGLITSTVQLVTQMLPYGCLLDHVRENRLIG 824
Db 757 TGPKANVFMDALIMASMDHPLVRLVGLVCLSPITQLVTQMLPHGCLLEYVHEKDNIG 816
Qy 825 SDDLNLWCQAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVH 884
Db 817 SOLLNLWCQAKGMVLEERLVHRDLAARNVLKSPNHVKITDFGLARLLGEGEKEYN 876
Qy 885 ADGKVKPKWMALESILRRRFTHQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKG 944
Db 877 ADGKVKPKWMALECIHVRKFTHQSDVMSYGVVWELMTFGKPYDGIIPAREIPDLLEKG 936
Qy 945 ERLPQPPCTIDVYMWKCMIDSECPRELVSFBRMARDPQRFVWQNEQ-LGPA 1003
Db 937 ERLPQPPCTIDVYMWKCMIDADSRPKFELAAEFBRMARDPQRYLVIQGDRLMKLP 996
Qy 1004 SPLDSTFRSLDEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHHRSSSRSGG 1063
Db 997 SPNUSKFFQNLDEEDLMDMAEYLV-P-QAFNIPP-----IYTSRARIIDSNRS--- 1046
Qy 1064 DLTGLGPESEAPRS-----PLAP-SEGAGSDVFDG 1094
Db 1047 --EIGHSPPAVTPMSGNQFVYVROGGAEGSVSPYRAPTSTIPEAPVAQATAEIED 1104
Qy 1095 DLGCAAGKLOSPLTHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVN 1147
Db 1105 SCCNGTLRKPAHVPHQVSDSTQRYSDPTVPAPERSPRGELDEEGYTMPMRDKPKQEYLN 1164
Qy 1148 QPQVRPQPPSPREGPLPAAPAGATLERAKTILSPGKNVWDVFAFGGAVENPEYLTQ 1207
Db 1165 PV-----ENPFVSR-----KNGDLQ-----ALONPEYHNASN 1193
Qy 1208 GAAPQHPHPPA-----FSPAFDNLVYWDOP 1233
Db 1194 G-----PPKADEYVNEPLYLNTFANTLGAEYLNKNNLSMPEKAKKAFDNPVWNHSL 1247
Qy 1234 PERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1248 PPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID ERB4 RAT
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCBI_TaxID=10116;
OX 11;
RN 1;
SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes."
RL J. Biol. Chem. 273:10261-10269(1998).
RN 2;
SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system."
RL Neuron 6:691-704(1991).
RN 3;
SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
CC 1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENCIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC 1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC 1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC 1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALLAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC 1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC 1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; --
CC EMBL; U52531; AAC53051.1; --
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD0000001; Euk_pkinase; 1.
```



```

Db 1195 ----PPKAEDVYNEPLVINTFNALGNAYMKNLSLVPEKAKAFDPDITWHSLLPR 1250
QY 1237 GA--PPSTFKGPT-----AENPEYL 1255
Db 1251 STLQHPDYLVQESTYKVFYKQNGRIRPIVAENPEYL 1285

RESULT 8
ID XMRK_XIPMA STANDARD; PRT; 1167 AA.
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA Telling A., Robertson S.M., Schartl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
RT inducing Tu locus in Xiphophorus.";
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16891; CAA34770.2; -.
CC PIR; S06142; S06142.
CC HSSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC PRINTS; PR00109; TYRKINASE.
CC PRODOM; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 5.
CC SMART; SM00220; S_TKc; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

```

```

FT DOMAIN 26 642
FT TRANSMEM 643 665
FT DOMAIN 666 1167
FT NP_BIND 710 977
FT BINDING 716 724
FT ACT_SITE 743 743
FT DISULFID 835 835
FT DISULFID 199 204
FT DISULFID 220 228
FT DISULFID 224 236
FT DISULFID 237 245
FT DISULFID 241 253
FT DISULFID 256 265
FT DISULFID 269 296
FT DISULFID 300 311
FT DISULFID 315 330
FT DISULFID 333 337
FT DISULFID 504 513
FT DISULFID 508 521
FT DISULFID 524 533
FT DISULFID 537 553
FT DISULFID 556 569
FT DISULFID 560 577
FT DISULFID 593 615
FT DISULFID 618 626
FT DISULFID 622 634
FT CARBOHYD 114 114
FT CARBOHYD 144 144
FT CARBOHYD 201 201
FT CARBOHYD 356 356
FT CARBOHYD 365 365
FT CARBOHYD 398 398
FT CARBOHYD 417 417
FT CARBOHYD 501 501
FT CARBOHYD 576 576
FT CARBOHYD 621 621
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 39.3%; Score 2690.5; DB 1; Length 1167;
Best Local Similarity 45.1%; Pred. No. 4.8e-136;
Matches 576; Conservative 165; Mismatches 385; Indels 151; Gaps 30;

QY 4 AALCRWGLLLALLPQGAAS-----QVCTGDMKRLRPASPETILDMRLHLYGCGVQVGN 59
Db 8 AALQ--LLLVLISRCSTDPDRKVCQGTSNQMTM---LDNHYLKKKKYKSGCNVLEN 62
QY 60 LELYLPNNAISLFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDN 119
Db 63 LEITYTQENQDLSFLQISQEVGYVLIAMNEVSTIPLVNLRLIRGNLYEGNFTLLVMEN 122
QY 120 GDPLNNTTPTVGASPGGLRELRLSLTEILKGGVLIQRNPQLCYODTILWKDF-----NNF 175
Db 123 YQK-NPSSP--DVYQVGLKQLQSLNLTILSGGVKVSHPNLLCNVETIWMMDIVDKTSNP 179
QY 176 TVSFWRVPKVSASHLENRSRACHPCSPMKRCWGSESSDCQSILTRVCAGGC-ARCK 234
Db 180 TML--IPHAF-----ERQCKDHGCVNGSWAPGPGHCQKFTKLLCAEQCNRRCR 229
QY 235 GPLPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFSPMPNPEGR 294
Db 230 GPKPTDCCNEHCAGCTGPRATDCLACRDFNDGDKTCTCPPPKIYDIVSHQVNDPNIK 289
QY 295 YTEGASCVTACPNYVNSTDVGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLGNE 354
Db 290 YTFGAACVKCEPSNYVTE-GACVRSCSAGMLEVD-ENGKRSCKPCDGVCPKVDGIGIG 347
QY 355 HLREVRVAVTSANIOEFAGCKIFGSLAFPLPESFDGDPASNTAPLOEQVFPETLEEITG 414
Db 348 SLSNTIAVNSNIRSFNCTKINGDIIILNRNSFEGDPHYKIGTMDPEHLMLTWTVEITG 407
QY 415 YLYTSAMPDLSPLDSVFQNLQVIRGRIILHNGAYS-LTLQGLIGISWLGSLRSLRELGLAL 473

```

```
Db 408 YLVIMWMPENMTSLSFQNLLEIRGRTTFSGFVVFVQVRHLQWLGRLSLKEVSAGNVI 467
Qy 474 IHHNTHLCFVHTVPDOLFRNPHOALLHTANRPEDECVEGLACHOLCARGHCWGPGPTQ 533
Db 468 LKNTQLQRLVANTINRRRLFRSDDQSLYDART-----ENQTCNNECEDGCGWGPPTM 520
Qy 534 CYNCSOFLRGQBCVEECRVLQGLPREYVNNARHCLPCHPECCQPNQSGSVTCFGPEADQCVAC 593
Db 521 CVSCSLHVDGRGRCVASCNLLQGEPREAQVDGRVQCHQCEQLVQTDLSLTCYGGPANCSSK 580
Qy 594 AHYKDPFVCVAPCSGVKPDLSYMPPIWKPFDEGACQPCPINCSTHSCVDLDDKQCPAER 653
Db 581 AHFDQGPQICPRCPHGILGDGTL-IWKYADKRGQCQCHQNCQTCQSGPGLSGCRGD-I 638
Qy 654 ASPLTSIVSAVGVILLVVLGVVFGFLIKRRROOKIRKYTMRLQLQELVEPLTPSGAMP 713
Db 639 VSHSSLAUGLVGLLITVALLIVLLRRRIK-RKTRIRCLLQKELVEPLTPSGQAP 697
Qy 714 NOAQRILKTELKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEI 773
Db 698 NOAFLRIKTEFKDKORVLGSGAFGVYKGLWNPDCGENIRIPVAIKVLEATSPKVNQEV 757
Qy 774 LDEAYVAGVSPYVSRLIGLICLTSTVQLVTQLMPYGCLLDVRNHRGLGSDLLNWC 833
Db 758 LDEAYVAGVSDPHVCRLLGICLTSAVQLVTQLMPYGCLLDVRNHRGLGSDLLNWC 817
Qy 834 QIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPK 893
Db 818 QIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPK 877
Qy 894 WMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQP 953
Db 878 WMALESILQWTVTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIASVLENGERLPQP 937
Qy 954 TIDVYIMVCKMIDSECPRELVSERWARDQRFVWQNEQDLGPASPLDSTFYRS 1013
Db 938 TIEVYMIILCKMIDSECPRELVSERWARDQRFVWQNEQDLGPASPLDSTFYRS 994
Qy 1014 LLEDDDMGLVDAEYLVPOQGFCCPDAPAGAGVMVHRHRSSTSGGGDLTLGLPSE 1073
Db 995 LLSSDD--DVDADEVLLPYKRI-----NRQS-----1020
Qy 1074 EAPRSLPAPSGAGSDVFDGLGMGAAGLQSLPHDPSLPQRYSEDFTV-PLPSETD 1132
Db 1021 -----EPCIPPTGH-----PVRENSITLRNISDPTQNALEKDLQ 1055
Qy 1133 YVAPLTCSPQEVNQPVRQP-----PSPRE-----GLP-AARAGATLERAKTL 1179
Db 1056 H-----EYVNPQGETSSRLSDIYNPNYEDLTGCMGPVSLSSQEAETNFSRPEYL 1105
Qy 1180 SPKGKGVKDVFAFGGAVENPEYLTPOGGAAPQHPHPAFSPAFNLVYWDQDPPRGPAP 1239
Db 1106 NTNQNSL---PLVSSGSMDDPY---QAG-----YQAF-----LPQTGAL 1140
Qy 1240 PSTFKGTPTAENPEYLG 1256
Db 1141 TGNMGFLPAENLEVLG 1157
```

RESULT 9

```
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Ising W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Shoyab M., Miki T., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Whittay M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RC MEDLINE=93282822; PubMed=7685162;
RA Kato M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; -
CC EMBL; M34309; AAA35979.1; -
CC EMBL; M61953; AAB26935.1; -
CC PIR; A36223; A36223.
CC HSSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD0000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrcK; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
```


KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 552 565
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT VARSPLIC 141 183

FT VARSPLIC 184 1342
FT CONFLICT 560 560
FT CONFLICT 1064 1064
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 35.4%; Score 2427.5; DB 1; Length 1342;
Best Local Similarity 40.5%; Pred. No. 5.7e-122;
Matches 533; Conservative 192; Mismatches 457; Indels 135; Gaps 32;

QY 10 GLLALLPGAA--STQVCTGDMKRLRASPETHLDMLRHLVQGVQVGNLELTLYLPT 67
DB 11 GLLFSLARGSEVGNVQVCPFTGLSVTGDENQVOTLYKLYERCEVVMGNLEIVLTGH 70
QY 68 NASLSFLODIQEVQGYVLIANNQVQLRQLRIVRGTLQFEDNYALVLDNGDPLNNTT 127
DB 71 NADLSFLOWIREVTGYVLVANNEFTSLPLNLRVVRVQVYDGKFAIFW-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSITEILKGGVLTORNPOLCYQDTILWKDFNNFTVSWLRVPKVS 187
DB 126 ----NSSHALQRLRUTQTEILSGGVYIEKNDKLCHMDTDWRDVR-----DRDA 172
QY 188 ASHLENRRACHPSCPMCKGRGCMGSESDCOSLRTVTCAGGC-ARCKGPLPTDCCHQC 246
DB 173 EIVVKDNGRSCPPCHEVCKG-RWGPGEDECOTLTKTICAPQCNGHCFGPNPQCCHDEC 231
QY 247 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 306
DB 232 AGGCGSGPQDTCFACRHFNDGACVPRCPQPLVYNKLTFOLEPNPHTKYQYGGVGVASCP 291

QY 307 YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGMEHURVAVTSAN 366
DB 292 HNFV-VDQTSVCVRACPPDKMEVD-KNGLKMXCEPCGGLCPKACGCTGSG--SRFTQTVDSN 347
QY 367 IQEFAGCKKI FGSFLAFPESEFDGDPASNTAPLOPEOLOVFETLEEITGYLYISAWPDSL 426
DB 348 IDGFVNCTKILNLDPLITGLNGDPWHKIPALDPEKLNVTFRVREITGYLNIOSWPHMH 407
QY 427 DLSVFQNLQVIRGRIHNGAYS-LTLQGLGISWLGSLRSURELGSGLALIHNNTHLCFVHT 485
DB 408 NFSVFSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGRSLKEISAGRIYISANQOLCVHHS 467
QY 486 VPWDQLFRNPQOALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFURQ 544
DB 468 LNWTKVLRGPTTEERLDIKHNRPRDCVAREGKVCPLCSCSGCGWGPQGLSCRNYSRG 527
QY 545 ECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPPECVA 604
DB 528 VCVTHCNFLNGEPREFAEAEFCSCHEPCQMEGTATNGSGSDTCAQCAHFRDGHCVS 587
QY 605 RCPGSGVPLDSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQRA----SPLTSI 660
DB 588 SCPHGVLG--AKGPIKYKYPDVQNECRPCHENCTQCGKPELODCLQTLVLKTHLTMA 645
QY 661 VSAVVGILLVVVLGVVFGILIKRQOKIR-KVTMRLLQETELVEPLTTPSGAMPNOAQM 719
DB 646 LTVIAG--LWIFMMLGGTFLYWRGRRIQNKAMRRYLERGESIEPLDPS-ERANKVYLAR 702
QY 720 ILKETELRKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPRANKIILDEAYV 779
DB 703 IFKETELRKVLGSGVGTGVHKGWVPEGESIKIPVCIKVIDKSGRGSFOAVTDHMLA 762
QY 780 MAGVSPVYSRLIGLCTSTVOLVTQLMFYGLLDHVRNRRGLSGQDLINMCQIAKGM 839
DB 763 IGSLDHAHIVRLGLCPGSSLVQYLPGLSLLDHVRQHRGALGQPLLNLNMGVQIAKGM 822
QY 840 SYLEDVRLVHRDLAARNVLVKS PNHVKITDPLGLARLLDIDETEHADGKVPKIMWALES 899
DB 823 YYLEEHGMVHRNLAARNVLLKSPQVQVADFGVADLLPPDDKQLLYSEAKTPIKWALES 882
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPOLLEKGERLQPPCTIDVYM 959
DB 893 IHFGKITQSDVMSYGVTVWELMTFGAEYAGLRLAEVDPDLEKGERLAQPOICTIDVYM 942
QY 960 IMVKCMIDSECRPRELVSFSESRMARDPQRFVWIONEDLGA----SPLDSTFYRSLLE 1016
DB 943 VMVKCMIDENIRPTFKELANEFTNRWADPPRYLVIKRES-GPGIAPGPEPHCLTNKKLE 1001
QY 1017 DDMGDVLDAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLEP-SEEE 1075
DB 1002 EVELEPELDLDLDEAED-----NLATTTLSGALSPLVGTNLRPR 1042
QY 1076 APRSPLAPSEAGSDVFDGDLGMAAGKLOSLPTH-D-PSPLORYSDPTVPLP-----S 1128
DB 1043 SQSLLSPSSGY-MPMNQNLGESCQESAVSGSSERCPVSLH-----PMPRGCLASE 1095
QY 1129 ETDGYVA-----PLTCSPOPE-----YVNPQDVRPOPSPREP----- 1162
DB 1096 SGEHVTGSEAELOEKVSMCRSRSRSPRPRGDSAYHSQRSHLLTPVTPSPGLEBED 1155
QY 1163 -----LPAARPAGATLERAKTLSP-GKNGW-----KDVFAFGGAVENPEYLTPOGGAA 1210
DB 1156 VNGYVMPDTHLKTGTPSSREGTLSSVGLSVLGTDEEDED-----EEYEMNRRRRHS 1207
QY 1211 POPHPPPAFSFAFDNLYYWD-----ODPPERCAPPSTFKGTAEENPEYL 1255
DB 1208 P-PHPHPPSSLEELGYEYMDVGSGLSASLGSTQSCPLHPVPIIMPATGTTPDEDEYEM 1263

RESULT 10
ERR3 RAT
ID ERR3 RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein";
RL Gene 165:279-284(1995).
RN (2)
RN REVISIONS TO 85; 513 AND 565.
RP Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29339; AAC28498.2; -;
CC EMBL; U52530; AAC53050.1; -;
CC HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transphosphatase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19 POTENTIAL
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	644	662	POTENTIAL.
FT	DOMAIN	663	1339	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	183	259	CYS-RICH.
FT	DOMAIN	707	964	PROTEIN KINASE.
FT	NP_BIND	713	721	ATP (BY SIMILARITY).
FT	BINDING	740	740	ATP (BY SIMILARITY).
FT	ACT_SITE	832	832	BY SIMILARITY.
FT	DISULFID	186	194	BY SIMILARITY.
FT	DISULFID	190	202	BY SIMILARITY.
FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AA5F2402BBDF1E CRC64;

Query Match 34.5%; Score 2364.5; DB 1; Length 1339;
Best Local Similarity 40.7%; Pred. No. 1,3e-118;
Matches 526; Conservative 174; Mismatches 426; Indels 165; Gaps 35;

Qy	3	LAALCRWGLLLALLPPGAA---	STQVCTGTDMLRLPASPEHLDMLRHLYQGCQVQGN	59
Db <td>7</td> <td>LQVLC----<td>FLISLARGSEMGNSQAVCPGTGLNGLSVTGDADNQYQTLKYKECEVWGN</td><td>62</td></td>	7	LQVLC---- <td>FLISLARGSEMGNSQAVCPGTGLNGLSVTGDADNQYQTLKYKECEVWGN</td> <td>62</td>	FLISLARGSEMGNSQAVCPGTGLNGLSVTGDADNQYQTLKYKECEVWGN	62
Qy <td>60</td> <td>LELTYLPTNASLFLQDIQEVQGYVLIAHNVQVPLQRLRIVRGTLQFEDNYALVDN</td> <td>119</td> <td></td>	60	LELTYLPTNASLFLQDIQEVQGYVLIAHNVQVPLQRLRIVRGTLQFEDNYALVDN	119	
Db <td>63</td> <td>LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLNLRVVRGTQVYDGKPAFVM--</td> <td>120</td> <td></td>	63	LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLNLRVVRGTQVYDGKPAFVM--	120	
Qy <td>120</td> <td>GDP LNNTTPVTGASPGGLRELQRLSLEILKGGVLIQPNQPCYQDTILWKDFNFTVSF</td> <td>179</td> <td></td>	120	GDP LNNTTPVTGASPGGLRELQRLSLEILKGGVLIQPNQPCYQDTILWKDFNFTVSF	179	
Db <td>121</td> <td>---LNYNT---NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCHMDITDRD-----</td> <td>165</td> <td></td>	121	---LNYNT---NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCHMDITDRD-----	165	
Qy <td>180</td> <td>WLRVPKVSASHL--ENRSRACHPCSPMGKSGRCWGESSEDCQSLTRTVACAGC--ARCKGP</td> <td>236</td> <td></td>	180	WLRVPKVSASHL--ENRSRACHPCSPMGKSGRCWGESSEDCQSLTRTVACAGC--ARCKGP	236	
Db <td>166</td> <td>---IVRVGAEIVVKNNGANCPCPCHEVCCKG--RCWGGPDDCQLITKTCAPQCNRCFCGP</td> <td>221</td> <td></td>	166	---IVRVGAEIVVKNNGANCPCPCHEVCCKG--RCWGGPDDCQLITKTCAPQCNRCFCGP	221	
Qy <td>237</td> <td>LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYT</td> <td>296</td> <td></td>	237	LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYT	296	
Db <td>222</td> <td>NPQCCHDECAGCGSPQDTCDFACRRFNDSGACVPCPEPLVYNKLTFLQLEPNPHTKYQ</td> <td>281</td> <td></td>	222	NPQCCHDECAGCGSPQDTCDFACRRFNDSGACVPCPEPLVYNKLTFLQLEPNPHTKYQ	281	
Qy <td>297</td> <td>FGASCVTACPNYVLTSDVSGCTLVCPHINQEVTAEDGTQRCCKSPCARVCYGLGMEHL</td> <td>356</td> <td></td>	297	FGASCVTACPNYVLTSDVSGCTLVCPHINQEVTAEDGTQRCCKSPCARVCYGLGMEHL	356	
Db <td>282</td> <td>YGGVCVASCAPHNFV--VDQTFQVACRACPPDKMEVD--KHGLKMCPCGGGLCKPRACEGTSG--</td> <td>337</td> <td></td>	282	YGGVCVASCAPHNFV--VDQTFQVACRACPPDKMEVD--KHGLKMCPCGGGLCKPRACEGTSG--	337	
Qy <td>357</td> <td>REVRVTSANIQEFACKIFGSLAFLPESFDCGDRASNTAPLOPEQLQVETLEEITGYL</td> <td>416</td> <td></td>	357	REVRVTSANIQEFACKIFGSLAFLPESFDCGDRASNTAPLOPEQLQVETLEEITGYL	416	
Db <td>338</td> <td>SRYQTVDSSNIDGFMVNTKILGNLDFLITGLNVDPMHKIPALDPEKLVNVRTVREITGYL</td> <td>397</td> <td></td>	338	SRYQTVDSSNIDGFMVNTKILGNLDFLITGLNVDPMHKIPALDPEKLVNVRTVREITGYL	397	
Qy <td>417</td> <td>YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYS--LTQLGLISWLGRLSRSLRELGLALIH</td> <td>475</td> <td></td>	417	YISAWPDSLPLDLSVFQNLQVIRGRILHNGAYS--LTQLGLISWLGRLSRSLRELGLALIH	475	

Db 398 NIOSPPHMHNFVSFNLTIGRSLYNRGFSLLIMKNLVNLSLGRSLKSEISARVYIS 457
Qy 476 HNTHLCFVHTVPMWDLFRNPHQALLHTA-NREDECVGEGLAHCHOLCARGHGWGPPTQC 534
Db 458 ANQOLCYHSLNWNTRLRGSPERLDIKYDRPLGCELAEGKVCDDLCSGGCGWGPQGQC 517
Qy 535 VNCOSQFLGOECVEECRVLOGLPREVNARHCLPCHPEQOPONGSVTCFPGPADOCVACA 594
Db 518 LSCRNYSREGVCTHCFNLQGPREFVHQAQCFSCHEPLMEGTSTCNGSGSDACARCA 577
Qy 595 HYKDPFFCVARCPGSKVDPDLSYMPIWKFPDEGACQPCPINCTHSC--VDLDDKGCAPAO 652
Db 578 HFDGPHCVNSCPHGLG--AKGPYKYPDQAQNECPCHENTCQCGPELQDCLQGAQV 635
Qy 653 RASPLTISVAVGVLLVVLGVVGLIKRQOKIR-KYTWRRLLQETELVEPIPTPSGA 711
Db 636 LMSKPHLVIAVTVG--LAVILMILGGSFLYMRGRIRKRAMRYLGERGESIEPLDPS-E 692
Qy 712 MENOAMRILKETELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNRENTSPKANK 771
Db 693 KANKVLARIFRETELRKVLKVLGSGVGTGKHGWIPEGESIKIPVCIKVIEDKSGRQSFQ 752
Qy 772 EILDVAVYAGVSGPYVSRLLGICLTSTVQLVTLMPYGCLLDHYRNRGRGLGSODLLNW 831
Db 753 AVTDHMLAVGSLDHAHIVRLGLCPGSSLOLVTLPLGSLLDHVKHRETLPQULLNW 812
Qy 832 CWOIAKGSYLEDVRLVRDLAARNVLKSPHNHVKITDPLGLARLLDIDETEHADGKVP 891
Db 813 GVOIAKGVYLEHSMVHRDLARNVLMKSPSQVQVADFGVADLLPPDDKQLLSEAKTP 872
Qy 892 IKWMALESILRRFTHQSDVNSYGVTVMLMTFGAKPYDGIPIAREIPDLLEKGERLPQP 951
Db 873 IKWMALESIHFGKYTHQSDVNSYGVTVMLMTFGAEPYAGURLAIPDLLEKGERLAQ 932
Qy 952 ICTIDVYIMVKCMIDSECRPRFRELSEFSRMARDPQRFVVIQNEIDLGPASLDSTFY 1011
Db 933 ICTIDVYIMVKCMIDENIRFTFKELANEFTRMARDPRYLVIKEAS-GFCTP--PAEE 989
Qy 1012 RSLLEDDMGDLVDAEYVLPQGFCDPDAPGAGMVHRRHSSTRSGGSDLTLGLEP 1071
Db 990 PSLVTTKEL-----QEAELEPEL-----DLDDLEA 1015
Qy 1072 SEE-----EAPRSLAPSEG-----AGSDVFDGLGWAAGK 1103
Db 1016 EEEGLATSLGALSILPTGTLTRPGSQSLSPSGYMPNMOSSLGEACLDASVAGREQF 1075
Qy 1104 LQSLPHTDPSLQRYSEDTVPLPSETGYV-----APL-----TC-----SPOPE--- 1144
Db 1076 SRPISLH-PIPRG-----PASESSEGHVGTGSAELQEKVSVCRSRSRSPRGRDS 1127
Qy 1145 -YVNPQDVRPOPSPSPREG-----LPAARPAGATILERAKTLSLPG-KNGYV----- 1187
Db 1128 AYHSORHSLTLPVTPSLSPGLEEDGNGYVMPDTHLRGASSSREGTLSSVGLSSVLGTEE 1187
Qy 1188 --KDVFAFGGAVENPEYLTPOGGAPOPHPP 1216
Db 1188 EDED-----EYEYMNRRRGSP-PRPP 1209

RESULT 11

EGFR_DROME STANDARD: PRT: 1426 AA.
AC P04412: O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DB Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpado protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CGL0079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
MEDLINE=94350209; PubMed=8070664;
Clifford R., Schubach T.;
"Molecular analysis of the Drosophila EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein.";
Genetics 137:531-550(1994).
[2]
REVIEWS.
Clifford R., Schubach T.;
Submitted (MAR:1998) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
MEDLINE=85124611; PubMed=2982499;
Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
"The Drosophila EGF receptor gene homolog: conservation of both hormone binding and kinase domains.";
Cell 40:599-607(1985).
[4]
SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
STRAIN=Oregon-R; Tissue=Embryo;
MEDLINE=87002474; PubMed=3093080;
Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
"Alternative 5' exons and tissue-specific expression of the Drosophila EGF receptor homolog transcripts.";
Cell 46:1091-1101(1986).
[5]
SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
MEDLINE=99102120; PubMed=9882502;
Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
"Several levels of EGF receptor signaling during photoreceptor specification in wild-type Ellipse, and null mutant Drosophila.";
Dev. Biol. 205:129-144(1999).
[6]
SEQUENCE FROM N.A. (ISOFORM TYPE I).
STRAIN=Berkley;
MEDLINE=20196006; PubMed=10731132;
Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Cocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Efannkoch C., Baldwin D., Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotter P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasseman D.A., Weinstock G.M., Weissbach J.,

Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
"The genome sequence of *Drosophila melanogaster*."
Science 287:2185-2195(2000).
[7]
SEQUENCE OF 959-1078 FROM N.A.
RP STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
growth factor receptor."
RL Nature 314:178-180(1985).
[8]
SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among *DER/fib* alleles: implications for
the mechanism of signal transduction by receptor-tyrosine kinases."
RL Genetics 129:191-201(1991).
[9]
RP REVIEW
RX MEDLINE=97248481; PubMed=9094709;
RA Perkinson N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
EGF receptor."
RL Cell 89:13-16(1997).
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF052754; AAC08536.1; -;
CC EMBL; AF052753; AAC08536.1; JOINED.
CC EMBL; AF052754; AAC08535.1; -;
CC EMBL; AF052752; AAC08535.1; JOINED.
CC EMBL; K03054; AAA51462.1; -;
CC EMBL; K03417; AAA51460.1; -;
CC EMBL; K03416; AAA50965.1; -;
CC EMBL; K03418; AAA51461.1; -;
CC EMBL; AF109077; AAD26134.1; -;
CC EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; -;
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; -;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -;
DR EMBL; AE003454; AAF46732.1; -;
DR EMBL; X02293; CAA26157.1; -;
DR EMBL; X78920; CAA55523.1; -;
DR EMBL; X78918; CAA55521.1; -;
DR EMBL; X78919; CAA55522.1; -;
DR PIR; A00640; GQFFE.
DR HSSP; P11362; IFGK.
DR FlyBase; FBgn003731; Egrf.
DR InterPro; IPR000494; EGR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PR00109; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.7%; Score 1964; DB 1; Length 1426;
Best Local Similarity 32.8%; Pred. No. 3e-97;
Matches 470; Conservative 190; Mismatches 422; Indels 352; Gaps 44;

QY 24 QVCTGDMKRLPASPETHLDMLRHLYGQCQVVOGNLELTYP-T-NASLSFLQDIOEVQG 82
DB 100 KICIGTKSRLSPVSKHHRNLRDRYNTCTVVDGNLKLTLPLNENLDLSFLDNIREVTG 159
QY 83 YVLIHNVQVQVLPQRLRIVRGTLF-----EDNVALAVLDNGDPLNNTPTVTGASPGGL 137
DB 160 YLILSHVDVKKVFPKLQIIRGTLFSLSVVEEKALFV-----TYSKM 203
QY 138 RELQLRSLEILKGVLIQNPQLCVQDTILWKDF--NNFTVSFW---LRVPKVSASHLEN 193
DB 204 YLEIPDLRLVNGQVGFHNNYLNCHMRTIQSEIVSNGTDAYNYVDFTAPE----- 255
QY 194 RSRACHPCSPMKGRCSRCSESSDQSLTRTRVCAGGCA--RCKGPLPTOCCHQCAAGCT 251
DB 256 --RECPKCHESCTHG--CWGEGPKNCQKFKSLTCSQCAGGRCYGPKPRECCHLFCAGGCT 312
QY 252 GPKHSDCLACLFHNSGICELHCPALVTYNTDTFFSMPNPEGRYTFGASCVTACPNYVLS 311
DB 313 GPTQKDCIACKNFFDEAVSKCECPMRKYNPTTYVLETNPEGKYAYGATCVKECP--GHLL 371
QY 312 TDVGSCTLVCPHLNQEVTAEDGTQCEKSKCARVCYGLGHEHLREVRVAVTSANTQFEA 371
DB 372 RONGACVRSRCPQDKMDKQGE-----CVPNGCPKPTCGVTVLH-----AGNIDSR 418
QY 372 GCKKIFGSLAFLPESFDG--DPASNTA-----PLQPEQLQVFTLEITGLYLSAWPD 423

Db 419 NCTVDCNIRILDTQSGFQDVYVMTGPRYIPLDPERREVSTVKEITGYLNIESTHP 478
Qy 424 SLPLDSVFNQLQVIRGRIHLNGAY-SLTQGLGISWLGRLSLRELGSGLALIHNNHLCF 482
Db 479 QFRNLVSFNLETHIRQLMESFAALAVKSSLSYLEMRNLKLISSGSGVVIQHRNDLCV 538
Qy 483 VHTVPMQDLFRNPHQALLTANRPEDECVEGGLACHQLCARGHCWGPPTQCVNCSQFLR 542
Db 539 VSNIRWPAIOLKEPEQKQVWVNEURLADLCEKNGTICSDQNEGCGAGTQCLTCQNF 598
Qy 543 GOECVEECVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC 602
Db 599 NGTCTIACGVISNAYK--PDNRCKICHPECR-----TCNGAGADHCQCVHVRDQHC 650
Qy 603 VARCP-----SGVK----- 611
Db 651 VSECPKNYKNGVRCRECHATCGCTGPKDTTIGACTTCNLAIINNDATVKRCLLKDK 710
Qy 612 -PLDSYMPWKF--PDEEGACOP-----CPI-----NCTH----- 638
Db 711 CPD-GY--FWEYVHQEQSLKPLAGRAVCRKCHPLCELCCTNYGVHQVCSKCTHYKRE 767
Qy 639 -----SC-----VDLDDKG----- 647
Db 768 QCETECPADHYTDEQRECFQRHPCNGCTGPGADDCKSRNFKLPDANETGPNVNTMF 827
Qy 648 -----CPAEOR-----ASPLTS-----IVSAVVGILLVVVLGV 675
Db 828 NCTSKCPLEMRHNVQYTAIGPYCAASPPRSSKITANLDVNMFIITGAVLVPTICILCV 887
Qy 676 VEGILIKRQOKIRKYT--MRLLQETELVEPLTPSGAMPNOAOMRIKTELKRVKVLG 733
Db 888 V--TVICROKAKETVKMTALSCDESEPLRPSNIGANLCUKRIVKDAELRGVJG 945
Qy 734 SGAGTVYKGIWIPGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVSRLLG 793
Db 946 MGAFRGVYKGVWVPEGENVKIPVAIKELKSTGAESSEBEFLREAYIMASEEHVLLKULA 1005
Qy 794 ICLTSTVOLVTLQMPYGLLDHVRNRLGSLQDLNLCWQIAGKMSYLEDVRLVHRDLA 853
Db 1006 VCMSSQMLITQLMPLGLCLLDVYRNRRDKIGSKALLNMTQIAGKMSYLEEKRLVHRDLA 1065
Qy 854 ARNLVKSPPNHKITTDFGLARLLDIDETEHADGGKVPKIKWMALESILRRRTHOSDYS 913
Db 1066 ARNLVQTPSLVKITDFGLAKLLSDSNEYKAAGGMPIKWLALCEIRNRVFTSKSDVMA 1125
Qy 914 YGVTVWELMTGAKPYDCIPAREIPDLLEKGERLPQPPICITIDVYMMVWKWMDSECRP 973
Db 1126 FGVTTIWELLTFQGRPHENIPAKDIPLDIEVGLKLEQPEICSLDIYCTLLSCHWLLDANRP 1185
Qy 974 RFRELVSFSRMDPQRFVVIONEDLG--PASPLDSTFYRSLLED--DMGDLVDABE 1028
Db 1186 TFKQLTTVFAEPDPRGYLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPTDDGSE 1238
Qy 1029 YLVPOQGFCDPAPGACGAVHHRSSSTRSGGDLTLGLEPSEEP-----RSPLA 1082
Db 1239 AIARPDYLYQKAAEGPS-----HRTDCT-----DEMPKLNRYCKDPSN 1277
Qy 1083 PSEGAGSVDFG---DLGNGAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTC 1139
Db 1278 KNSSTGDDERDSSAREVGVGNLR-----LDLPVDEDDLYMP-TC 1315
Qy 1140 SPOEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVPFAGGAVEN 1199
Db 1316 QPGNNNNMMN-----NPNQNNMAAVGAAGYM-----DLIGVPVSVDN 1354
Qy 1200 PEYL-----TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1230
Db 1355 PEYLLNAQTLVGESPIQTQIGIPVMGPGTMEVKVMPGSEPTSSDHEYND 1408

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]_SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10056; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glucocorticoid; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.5%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. NO. 3.1e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 593 CAHYKDPFPCVARGSPGVKPDLSYMPKPFDEEGACQPCPINCTHSCVDLDDKCPARQ 652
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPCNTRCGKPGLEGCP--- 58
Qy 653 RASPLTSIVSAV--GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGA 711
Db 59 NGSKTPSIAAGVGGELCLVVVGLGIGLYLARR-HIVKRTLRLLQRELVEPLTPSGE 117
Qy 712 MPNQAMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPKAN 771

```
Db 118 APNQHRLRLKETFEKVKVVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDEAYNAGVSPVSRLLGICLTSTVOLTPMLPYGCLLDHVRNRRGLSGSODLLNW 831
Db 178 EILDEAYNWSVDNPHVCRLLGICLTSTVOLTPMLPYGCLLDYIHKHNDTNGSQVLLNW 237
Qy 832 CMQIAGMSYLEDLVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVP 891
Db 238 CVQIAGMNYLEERLVRDLAARNVLVKTPOHVKITDFGLAKLLGADKEVHAEGCKVP 297
Qy 892 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSEKWARDPPRYLVIOGDERMHLPSPTDSKF 417
Qy 1011 YRSLLDEDDMDGLVDAEYLVPOQGFCDPPAPGAGMVMHRRSSSTRSGGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRSL-----APSEGAGSDVFDGLGMAAKGLQSLTHDPSPLOYSEDPVP 1125
Db 450 -----SRTPLLSSUSATSNNSATNCID-----RNGQGHVREDTSFVQRYSSDPTGN 495
Qy 1126 LPSET--DGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLPSPCK 1183
Db 496 FLEESIDDDGL-----PAPEYVQ--LMPKPS-----TAMVQ 526
Qy 1184 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLY 1228
Db 527 NOIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1229 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 579 WTQSGNHQINLONPDYQDQFLNPETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
(1)
RN SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN (2)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
```

```
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSP; P11362; LFQK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
FT SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;
Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.7e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
Qy 593 CAHYKDPFCVACPSGVKPDLSYMPKFPDEEGACQPCINCHSCVDLDDKGPAG 652
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VRKYADANAVCQLCHPNCCTGCKGPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGA 711
Db 59 NSGKTPSIAAGVVGGLLVVGLGVLRR-HIVRKRTLRLQLQELVEPLTPSGE 117
Qy 712 MPNQMRILKETELRKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 771
Db 118 APNQHRLRLKETFEKVKVVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 772 EILDEAYNAGVSPVSRLLGICLTSTVOLTPMLPYGCLLDHVRNRRGLSGSODLLNW 831
Db 178 EILDEAYNWSVDNPHVCRLLGICLTSTVOLTPMLPYGCLLDYIHKHNDTNGSQVLLNW 237
Qy 832 CMQIAGMSYLEDLVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVP 891
Db 238 CVQIAGMNYLEERLVRDLAARNVLVKTPOHVKITDFGLAKLLGADKEVHAEGCKVP 297
Qy 892 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSEKWARDPPRYLVIOGDERMHLPSPTDSKF 417
Qy 1011 YRSLLDEDDMDGLVDAEYLVPOQGFCDPPAPGAGMVMHRRSSSTRSGGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRSL-----APSEGAGSDVFDGLGMAAKGLQSLTHDPSPLOYSEDPVP 1125
```


CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	ENBL; M20386; AAA48760.1; -.
DR	InterPro; IPR000494; EGFR_L_domain.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR002174; Furin-like.
DR	InterPro; IPR001245; Tyr_pkinase.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L_domain; 2.
DR	SMART; SMO0261; FU; 4.
DR	DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR	DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR	DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; PARTIAL.
KW	Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW	Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT	SIGNAL 1 30
FT	CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT	DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 655 667 POTENTIAL.
FT	DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT	DISULFID 197 206 BY SIMILARITY.
FT	DISULFID 201 214 BY SIMILARITY.
FT	DISULFID 222 230 BY SIMILARITY.
FT	DISULFID 226 238 BY SIMILARITY.
FT	DISULFID 239 247 BY SIMILARITY.
FT	DISULFID 243 255 BY SIMILARITY.
FT	DISULFID 258 267 BY SIMILARITY.
FT	DISULFID 271 298 BY SIMILARITY.
FT	DISULFID 302 314 BY SIMILARITY.
FT	DISULFID 318 333 BY SIMILARITY.
FT	DISULFID 336 340 BY SIMILARITY.
FT	DISULFID 513 522 BY SIMILARITY.
FT	DISULFID 517 530 BY SIMILARITY.
FT	DISULFID 533 542 BY SIMILARITY.
FT	DISULFID 546 562 BY SIMILARITY.
FT	DISULFID 565 581 BY SIMILARITY.
FT	DISULFID 569 589 BY SIMILARITY.
FT	DISULFID 592 601 BY SIMILARITY.
FT	DISULFID 605 627 BY SIMILARITY.
FT	DISULFID 630 638 BY SIMILARITY.
FT	DISULFID 634 646 BY SIMILARITY.
FT	CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	NON_TER 703 703 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

```
Query Match      23.5%; Score 1611; DB 1; Length 703;
Best Local Similarity 44.4%; Pred. No. 8.2e-79;
Matches * 316; Conservative 110; Mismatches 254; Indels 32; Gaps 12;

QY      8 RWGLLLALLPPGAA-----STVCVCTGDKMLRLPASPETHLDMLRHLYGGCCVVQGNLE 61
          :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
          :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
```

```

13  RGAUVLVLLGLVALCSAUEKKVCOGTNNKLTQLGHVEDHFTSLQRMYNCEVVLNSLE  72
62  LTYLPTNASLFLQDIOEVOGYVLIAHNQVRQVQLRLRIVRGTLQFEDNYALAVLDNGD  121
73  ITVYEHNRDLTFLKTIQEVAGYVLIALNMVDVI PLENLQITIRGNVLVDNSFALVLSNYH  132
122  PLANNTTPVTGASPGGLRELOLRSLTEILKGGVILQORNPOLCYODTILKDKFNNTVSPWL  181
133  -MNKQT-----GLRELPMKRUSEITLNGVKISNNPKLCNMOTVLWNDDIITS-----  178
182  RVPKVSASHLENRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVVCAGGCA-RCXGKPLPTD  240
179  RKPLTVLDFASNL5-SCPCKHPNCTEDHCWGAGEQNCOTLTKVICAQOCSGRCRKQVPSD  237
241  CCHQCAAGCTGPKHSOCLACLHNHSGIELHCPALVTYNTDTFESMPNPEGTYTFGAS  300
238  CCHNQCAAGCTGPRSEDCACRKFRRDATCKDTCPPLVLYNPTTYQMDVNPGEKYSFGAT  297
301  CVTACPNYLSTDVSGSTCLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR  360
298  CVRECPHNVYVTDHGSCVRSQNTDITYEV-FENGVRKCKKCDGLSKSKVNGIGIELKGIL  356
361  AVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQOVFETLEEITGYLYISA  420
357  SINATNIDSFNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFTVKEISGFLLIOA  416
421  WPD5LPLDVFNQLQVIRGIRLHNGAYSILTLOGLGISWLGRLSRLRELGSLGLALHHNTHL  480
417  WPDNATLYAFENLEIIRGTRTKQHGQYSVLAVNLIKIOSLGRLSUKETSDGDIAMKNKNL  476
481  CFVHTVPWDQLFRNPQHALLHTANRPEDECVEGELACHQLCARGHCGWPGPTQCVNCSQF  540
477  CYADTNWRSIFATOSQKTKIIQNRKNNDCTADRHVCDPLDCSDVCGWCGPGPFHCFSCRF  536
541  LRQGEVCNEBRLVLOGLPREVYNARHCLPCHPEQOPQNG----SVTCFPEADQCACVAHYK  597
537  SRQKECVKQCNILQGEPEPREFRSDKPLCHSECLVQNSTAYNTTCSGPGDCHMKCAHFI  596
598  DPPPCVARCP5GVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKQCPAEORASPL  657
597  DGPCHVCACPAVLGENDTL-VKYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKT  652
658  TSIVSAVV-GILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQOETLEVEPLTP  708
653  PSTAAGVVGGLCLVAVVGLGTLGYLRRR-HIVRKTRLRLLQOERELVEPLTP  703

```

Search completed: July 22, 2003, 08:44:56
Job time : 22.2793 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-149-163-14
Perfect score: 6848
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVP 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6198	90.5	1259	6 O18735	O18735 canis famil
2	3166	46.2	1209	11 Q9QX70	Q9QX70 rattus norv
3	3139	45.8	1210	11 Q9EP98	Q9EP98 mus musculu
4	2724	39.8	1165	13 Q9VH40	Q9VH40 xiphophorus
5	2711.5	39.6	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2307	33.7	1328	13 P73754	P73754 fugu rubrip
7	2047.5	29.9	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1777.5	26.0	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.4	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.1	729	15 Q8G712	Q8G712 avian rous-
11	1718	25.1	567	15 Q8G714	Q8G714 avian rous-
12	1697.5	24.8	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1653.5	24.1	962	15 Q64895	Q64895 avian eryth
14	1645	24.0	545	15 Q85468	Q85468 avian eryth
15	1530.5	22.3	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1514.5	22.1	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1273	18.6	1193	5 Q9Y1X8	Q9Y1X8 ephydatia f
18	1194.5	17.4	1368	5 Q23821	Q23821 caenorhabdi
19	1157	16.9	1717	5 Q26566	Q26566 schistosoma
20	1142	16.7	527	13 Q90836	Q90836 gallus gall
21	1025.5	15.0	478	11 Q9ESE0	Q9ESE0 rattus norv
22	952.5	13.9	599	13 Q9PSH2	Q9PSH2 gallus gall
23	906	13.2	165	4 Q14256	Q14256 homo sapien
24	887	13.0	176	11 Q923V5	Q923V5 rattus norv
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 Q8SZW1	Q8SZW1 drosophila
27	754.5	11.0	311	13 Q99162	Q99162 xiphophorus
28	749.5	10.9	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
29	725	10.6	1671	5 Q9NJV5	Q9NJV5 biomphalari
30	722	10.5	149	6 Q9BG66	Q9BG66 oryctolagus
31	721	10.5	331	4 Q9BUD7	Q9BUD7 homo sapien
32	702	10.3	1418	13 Q93457	Q93457 scophthalmu
33	699.5	10.2	1368	13 Q8UW85	Q8UW85 paralichthy
34	682	10.0	1358	13 Q71798	Q71798 xenopus lae
35	677.5	9.9	1369	13 Q8UW86	Q8UW86 paralichthy
36	665.5	9.7	1472	5 Q9U5A8	Q9U5A8 bombyx mori
37	660	9.6	1412	13 Q8UW84	Q8UW84 paralichthy
38	647.5	9.5	1418	13 Q8UW83	Q8UW83 paralichthy
39	642	9.4	1245	13 Q9YGH8	Q9YGH8 scophthalmu
40	639	9.3	1371	11 Q9QVW4	Q9QVW4 rattus sp.
41	598	8.7	987	11 Q9IYM0	Q9IYM0 mus musculu
42	597	8.7	935	4 Q96L35	Q96L35 mus musculu
43	595	8.7	987	11 Q99MR2	Q99MR2 mus musculu
44	594.5	8.7	2144	5 Q9VD94	Q9VD94 drosophila
45	587.5	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1
O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB008451; BAA23127.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00089; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TVR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D45C4ACD46 CRC64;

Query Match	90.5%;	Score 6198;	DB 6;	Length 1259;
Best Local Similarity	90.4%;	Pred. No. 0;		
Matches 1148;	Conservative 44;	Mismatches 58;	Indels 20;	Gaps 5;
Qy	1	MELAAALCRWGLLLALLPPGAAS	60	
Db	1	MELAAWCRWGLLLALLPPSAGT	60	
Qy	61	ELTYLPTNASLSFLQDQEVGYVLI	120	
Db	61	ELTYLPANASLSFLQDQEVGYVLI	120	
Qy	121	DPLNNTPTVGTASPGGLRELQRLS	177	
Db	121	DPLEGGIPAPGAAGGLRELQRLS	177	
Qy	178	SFWLRVPKVSASHLE-NRSRACHPC	236	
Db	178	-----QLAULTLIDTNFSAACPP	230	
Qy	237	LPTDCHEQCAAGCTGPKHSDCLAC	296	
Db	231	QPTDCHEQCAAGCTGPKHSDCLAC	290	
Qy	297	FGASCVTACPNYSLTDVSGSCTLV	356	
Db	291	FGASCVTSCPNYSLTDVSGSCTLV	350	
Qy	357	REVRANTSANTQEFAGCKKIFGSLA	416	
Db	351	REVRANTSANTQEFAGCKKIFGSLA	410	
Qy	417	YISAWPDSLPDLSVFQNLQVIRGR	476	
Db	411	YISAWPDSLPNLSVFQNLRVIRGR	470	
Qy	477	NTHLCFVHTVWDQLFRNPHQALLH	536	
Db	471	NARLCFVHTVWDQLFRNPHQALLH	529	
Qy	537	CSQFLRGCEVCECRVLQGLPREYV	596	
Db	530	CSQFLRGCEVCECRVLQGLPREYV	589	
Qy	597	KDPFPCVACRCSGVKPDLSYMPIW	656	
Db	590	KDPFPCVACRCSGVKPDLSFPMKFA	649	
Qy	657	LTSIVSAVVGILLVVVLGVVFGILI	716	
Db	650	VTSIIAAVVGILLAAVWGLVGLIKR	709	
Qy	717	QMRILKETELRKVKVLSGAFCTVYK	776	
Db	710	QMRILKETELRKVKVLSGAFCTVYK	769	
Qy	777	AYVMAGVSPVSRLLGICLSTVQLV	836	
Db	770	AYVMAGVSPVSRLLGICLSTVQLV	829	
Qy	837	KGMSYLEVRLVHRDLAARNVLVKS	896	
Db	830	KGMSYLEVRLVHRDLAARNVLVKS	889	
Qy	897	LESILRRFTHOSDWSVGVTVWELMT	956	
Db	890	LESIPPRFTHOSDWSVGVTVWELMT	949	
Qy	957	VYIMVWKCMIDSECRPRFRELVAE	1016	
Db	950	VYIMVWKCMIDSECRPRFRELVAE	1009	
Qy	1017	DDMDGLVDABEYLVPOQGFPCPD	1076	

Db	1010	DDMDGLVDABEYLVPOQGFPCPE	1069	
Qy	1077	PRSLAPSEAGSDVFDGLGMAAKGL	1136	
Db	1070	PKSLAPSEAGSDVFDGLGMAAKGL	1129	
Qy	1137	LTCSPQPEYVNVQDVRPQPSPREG	1191	
Db	1130	LTCSPQPEYVNVQDVRPQPSPREG	1189	
Qy	1192	AFGAVENPEYLTPOGGAAPQPHPP	1251	
Db	1190	AFGAVENPEYLTAPRGAAPQPHPP	1249	
Qy	1252	PEYGLDVPV	1261	
Db	1250	PEYGLDVPV	1259	

RESULT 2								
Q9QX70		PRELIMINARY;	PRT;	1209	AA.			
ID	Q9QX70							
DT	01-MAY-2000	(TrEMBLrel. 13, Created)						
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)						
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)						
DE	Epidermal growth factor receptor.							
GN	EGFR.							
OS	Rattus norvegicus (Rat).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.							
OX	NCBI_TaxID=10116;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=FISHER; TISSUE=LIVER;							
RX	MEDLINE=90258888; PubMed=2342466;							
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,							
RA	Earp H.S.;							
RT	"A truncated, secreted form of the epidermal growth factor receptor is							
RT	encoded by an alternatively spliced transcript in normal rat tissue.";							
RL	Mol. Cell. Biol. 10:2973-2982(1990).							
RN	[2]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=FISHER; TISSUE=LIVER;							
RA	Petch L.A.;							
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.							
RN	[3]							
RP	SEQUENCE FROM N.A.							
RC	STRAIN=FISHER; TISSUE=LIVER;							
RA	Guttridge K., Dawson T.L., Earp H.S.;							
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; M37394; AAF14008.1; -.							
DR	HSPF; P13362; IFGK.							
DR	InterPro; IPR000494; EGFR_L_domain.							
DR	InterPro; IPR000719; Euk_pkinase.							
DR	InterPro; IPR002174; Furin-like.							
DR	InterPro; IPR001245; Tyr_pkinase.							
DR	Pfam; PF00757; Furin-like; 1.							
DR	Pfam; PF00069; pkinase; 1.							
DR	Pfam; PF01030; Recep_L_domain; 2.							
DR	PRINTS; PR00109; TYRKINASE.							
DR	ProDom; PD000001; Euk_pkinase; 1.							
DR	SMART; SM00261; FU; 3.							
DR	SMART; SM00219; TyrKc; 1.							
DR	PROSITE; PS00107; PROTEIN KINASE_ATP; 1.							
DR	PROSITE; PS50011; PROTEIN KINASE_DOM; 1.							
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.							
KW	ATP-binding; Receptor; Tyrosine-protein kinase.							
SEQUENCE	1209	AA;	134891	MM;	96FEE7F6CC1B7773			
					CRC64;			

Query Match	46.2%;	Score 3166;	DB 11;	Length 1209;
Best Local Similarity	49.8%;	Pred. No. 5.2e-230;		

Matches 641; Conservative 166; Mismatches 353; Indels 126; Gaps 26;		
QY	3	LAALRCWGLJALLPPGA-ASTOVCTGDMKLRLPASPETHLDMRLHYLCQCVVQGNLE 61
DB	15	LAALCAAG-----GALEKKVCOGTSNRLTQGTGFEDHFLSLQRMFNCEVVLGNLE 66
QY	62	LTYPNTASLSFLQDIQEVQYVLIHNVQVPLQRLRIRVRGTQOLFEDNYVALVONGD 121
DB	67	ITYVQRYNDLSFLKTIOEVAGYVLIATNVERIPIENLIQIIRGNALYENTVALAVLSN-- 124
QY	122	PLNNTTPTVGTGPGSLRELQRLSLTEILKGGVLTORNPOLCYQDTILWKDPNNFTVSFWL 181
DB	125	-----YGNKTKGLRELPRNLEQELIIGAVRSNPILCNMETIOWROI----- 167
QY	182	RVPKYSASHL-----ENRSRACHPGSPCKGSRCKGESSEDCQSILTRTVCAAGCA-RCKGP 236
DB	168	-VDVFLSNMSMDVQRHLTGCPKCDPSPNGSCWGRGEENCOKLTKIIAQOCSRRCRGR 226
QY	237	LPTDCCHQCAAGCTGPKHSDCLACHFNHSGICELHCPALVTVNTDTFESMPNPEGRYT 296
DB	227	SPSCCHNQCAAGCTGPRSDCLVCHRDRDEATCKDTCPLMLYNPTTYQMDVNPPEGYS 286
QY	297	FGASVCTACPNYLSTDVSGCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356
DB	287	FGATCVKCPRYVTVTHGSCVCRAGDPDYEV-BEDGVSKCKCDGCRKVCNGIGEF 345
QY	357	REVRAVTSANIOEFAGCKKIPGSLAFIPESFDGDPASNTAPLOEQLOQVFETLEEITGYL 416
DB	346	KDTLSINATNIKFKYCTAISGDHLHPVAFKGDSTRTPLDPRELEILTKVKEITGFL 405
QY	417	YISAWPDSLPDLSVFONLOVIRGILHNGAYSLTLOGIGISWLGRLSRLRELGSGLAIHH 476
DB	406	LIQAWPENWTDLHAFENLEIIRGTOKHGFQSLAVGLNITSLGLRSLEISDGDVITISG 465
QY	477	NTHLCFVHTVPDOLFNRPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGRCPTOCVN 536
DB	466	NRNLCYANTINWKKLFTPNQTKIMNRAEKCKATNHCNPLCSSEGCGWGEPTDCVS 525
QY	537	CSQFLRQGECEBECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVCAHY 596
DB	526	QNVSRGRECVKCNILEGEPRFVENSECICQHPCELPQTMNITTCGRGPDNCIKCAHY 595
QY	597	KDPFCFCVACPSGVKPDLSYMTKFPDESGACOPCINCTHSCVDLDDKGCFAQRASP 656
DB	586	VDGPHCVKTCPSGIMGNNTL-VKFDADANNVCHLCHANCTYCAGFLKGC--QOPEGP 642
QY	657	-LTSIVSAVVGILLVVLGVVFGI-LIKRQOKIRKVTMRLLQETELVEPLTPSGAMPN 714
DB	643	KIPSIATGVGGLLFIVV-VALGIGLFWRRRLVQKRTLRLLQERELVEPLTPSGEAPN 701
QY	715	QAQNRILKTELKRVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPPKANKEIL 774
DB	702	QAHLRLIKTEFKIKVLGSGAFGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEIL 761
QY	775	DEAVMAGVGSPPVSRLLGLCLSTVOLVTLQMPYGLLDHVRNCRGLSGDQLLWNCMQ 834
DB	762	DEAVMASVDNPHVCRULGCLSTVOLITQMPYGLLDYVREHKDNIGSQYLLWNCVQ 821
QY	835	IAGKMSYLEDLVRLVHRLDAARNVLVSPNHVKITDFGLARLLDIDETEHADGCKVPIKW 894
DB	822	IAGKMYLEDRRLVHRLDAARNVLVTPQHKITDFGLAKLGAKEEYHAEKGKVPKIKW 881
QY	895	MALESILRRFTQSDVWSGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICT 954
DB	882	MALESILHRTYTHQSDVWSGVTVWELMTGSKPYDGIPIASEISSILEKGERLPQPPICT 941
QY	955	IDVYIMVWKWIMDSECRPRELVSFESRWARDQRFVVIQ-NEDLGPASPLDSFYRS 1013
DB	942	IDVYIMVWKWIMDADRSPKRELIEFLFSKQAKDPQRYLVLIQGDERNHLPSPDSNFYKA 1001
QY	1014	LLEDDMDGLVDAAEYLVPOQGFCCPDPAPGAGGVHRRHRSSTSGGDLTLGLEPSE 1073
DB	1002	LMEEDMEDVDADEVLIPOQGF-----NSPST----- 1030

Best Local Similarity 46.9%; Pred. No. 1e-195;		Matches 540; Conservative 170; Mismatches 346; Indels 95; Gaps 27;	
Qy	161	LCVQDTTLWKDF-----NNTVSFWLRVPKVSASHLENRSRACHPCSPCKGSRCKGES	214
Db	3	LCFADTTTHQDIVRNWASNFTL-----VP-----TNGSSGCGCRCKSKCTG-RCWGPT	49
Qy	215	SEDCOSLTRVCAGGC-ARCKGPLPTCCHEQCAAGCTGPKHSCLACLPHNHSIGICELH	273
Db	50	ENHCOTLTKVCAEQDGRCVGPVSDCHREAGCGGPKDTCFACMFNDSGACVQ	109
Qy	274	CPALVYNTDTFESMPNPEGRTFGASCVTACPNYLSTDVGSCTLCPLHNQEVTAEDG	333
Db	110	CPQTFVYNTPTFQLEHNHNAKYTGAFCKVKCPNFV-VDSSSCVRACPPSSKMEV-EE	167
Qy	334	TQRECKSKCARVCYGLGHEHLREVRAVTSANTQEPAGCKIIFGSLAFLPESFDGPAS	393
Db	168	IKMKCPCTDTCPRACDGGTGSLSVAQTVSSNDIKFINCTKINGNLIFLVTGTHGDPYH	227
Qy	394	NTAPLOEQLOVFTLSEITGVLVISAPDPSLPSVFNQVIRGRILHNGAYSLLTQ	453
Db	228	TIAAINEKLNIFQTVREITGVLNIQSWPNWTDVRFVSNLVTIGGRALVGLSLLILKQ	287
Qy	454	LGISWLGRSLRELGSGLALHNNHNLFCVHTVPMQDLFRNPQALLHTANRPEDEC	513
Db	288	QGITSLOFQSLKQISAGNIYITDNLNLCYYHTVNTSLFSTPSQKTVIHRNKAENCTAD	347
Qy	514	GLACHQICARGHCWGPPTQCVNCSQFLRGQECVEBCEVLOGLPREVNAHCLPCHPEC	573
Db	348	GMVNELCSSDGCWGPDOCLSCRFIRGTCIESCNLYDGEFREFANGSVNCECDPQC	407
Qy	574	QP-ONGSVTCFGEADOCVACAHYKDPFCVAPCPGVKPDLSYMPITWKPDERGACQC	632
Db	408	EKVEDNMITYCGPDPDCTCKCFHFKDGPNCVEKCPDGLQANSF--IFKVADEDRCHPC	465
Qy	633	PINTHSCVDLDDKGC-----PAEQRASPLTSIVSAVY-GILLVVLGVVFGL	680
Db	466	HPNCTQCGRPASHDCIYVWPTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVY	522
Qy	681	IKRQOKIRKYWRLLQETVELVEPLTPSGAMPNQAQWRIILKETELRKVKVLSGAGTV	740
Db	523	VRRSIK-KKRALRRFL-ETELVEPLTPSGAPNQAQRLILKETELRKVKVLSGAGTV	580
Qy	741	YKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVWAGVSGVSVRLGLCLSTV	800
Db	581	YKGIWVPEGETVKIPVAIKILNETTGPKANVEFNDEALIMASMDPHVRLVGLCSPTI	640
Qy	801	QLVTQLMPYGLLDHVRNRRGLSGQDLLNMCQIAKMSYLEVDRLVHRDLAARNVLK	860
Db	641	QLVTQLMPHGLLDYVHEHKDNIQSOLLNMCVQIAKGMYLEERLVRDLAARNVLK	700
Qy	861	SPNHVKITDGLARLLDIDETEXHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWE	920
Db	701	SPNHVKITDGLARLLSGDEKEYNADGGKVPKIMMALECIHYRKFTHQSDVMSYGVTVWE	760
Qy	921	LMTFGAPYDGIPIAREIPDLLEKGERLPQPPICTIDVYVMVWKCWMTDSECRPFRELV	980
Db	761	LMTFGGRPYDGIPIREIPDLLEKGERLPQPPICTIDVYVMVWKCWMTDASRPFKELAA	820
Qy	981	EFMRMARDPQRFVVIQNEED-LGPASPLDSTFYRSLLEDDDMGLVDAAEYLVPOQGF	1039
Db	821	EFMRMARDPQRYLVQIDDRMKLPSPNDSKFFQNLDEEDLEDMWDAEYLVLP-QAFNIP	879
Qy	1040	DPAPGAGMWHHRHSSTSGGGLTLGLEPSEEAPRS--PLAP-SEGAGSDVFDGLD	1096
Db	880	PIYTSRTRISNRNQFYRDGGVAABGVV-PMPYRAPGCIIPAPVAQGATAIFEDTC	938
Qy	1097	GMGAALKQLSLTHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQEVYNQ	1149
Db	939	CNGTLRKQVATLAKEDSSQRYSDPTVFIPERVIRGELDEDGYMTWRDKPKTDYLNVP	998
Qy	1150	DVRPQPPSPREGPLPAA-RPAGATLERAKTSLSPKNGVVKDVF-----AFGGAENP	1200
Db	999	EENPFVSRKXNGDLQAVDNPEYHN-----APNGQPKAEDEYVNEPLYLNTFANTLENA	1051
Qy	1201	EYLTPQGAAPQHPHPAPSPAFDNLYYWDODPPERGA--PPSTFKGTPT-----	1248
Db	1052	EYL-----KNNLPEKAKAFNDPDYMNHSLPPRSTLQHPDYLQBYSTKYFYKONGRI	1103
Qy	1249	-----AENPEYL 1255	
Db	1104	RPIVAENPEYL 1114	
RESULT 6			
ID	P79754	PRELIMINARY;	PRT; 1328 AA.
AC	P79754;		
DT	01-MAY-1997	(TrEMBLrel. 03, Created)	
..T	01-MAY-1997	(TrEMBLrel. 03, Last sequence update)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)	
DE	ExB83.		
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;		
OC	Tetraodontidae; Takifugu.		
OX	NCBI_TaxID=31033;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=99177347; PubMed=10077531;		
RX	Gelliner K., Brenner S.;		
RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu		
RT	rubripes";		
RL	Genome Res. 9:251-258(1999).		
DR	EMBL; AF056116; AAC34391.1; ..		
DR	HSSP; P11362; IFGK.		
DR	InterPro; IPR000494; EGFR_L_domain.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; pkinase; 1.		
DR	Pfam; PF01030; Recep_L_domain; 2.		
DR	ProDom; PD000001; Euk_pkinase; 1.		
DR	SMART; SM00261; FU; 3.		
DR	SMART; SM00219; TyKc; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.		
KW	ATP-binding; Transferase.		
SQ	SEQUENCE 1328 AA; 148613 MW; A3330392588647E9 CRC64;		
Query Match 33.7%; Score 2307; DB 13; Length 1328;			
Best Local Similarity 40.5%; Pred. No. 4.6e-165;			
Matches 525; Conservative 152; Mismatches 414; Indels 206; Gaps 33;			
Qy	9	WGULLALLPP--GAASTQ-----VCTGDMKRLRLPASPETHLDMLRHLHYQGCVVQGNLEL	62
Db	4	WRLLIMCVASRLRAASSQTQEAQVCPGTQNGLSSTGSEQNYLNKDKYKGEIIMGNLEI	63
Qy	63	TYLPTNASLFLQDIOEQVQVYLIAHNQVQVLPQRLRIRVGTOLFDENYALAVLDNGDP	122
Db	64	QTIESNWDFFSLKTIREVTGYVLIAMNHQEIPLGQLRVIRGNSLYERRFALSVELN---	120
Qy	123	LNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQNPQLCYQDTILWKDF--NN---FTV	177
Db	121	-----YPKDG--PSGLNQLGLMNLTEILDGGVQVINNKYLRYPGWVYVYRDIIRNDAPTEI	174
Qy	178	SFWLRVPKVSASHLENRSRACHPCSPMCKGSRCKGESDQCSLTRVCAGGC-ARCKGP	236
Db	175	QF-----NGERG--CHKSC-GNYCWGPGKQCCQLTKTVCAPOCNDRCFGT	218
Qy	237	LPTDCCHQCAAGCTGPKHSCLACLPHNHSIGICELHCPALVYNTDTFESMPNPEGRT	296
Db	219	SPRDCCHIECAAGCKGKPLDTCFACRLFNDSGACVPOCPQTLIYNKQTFQMETPNNAKYQ	278


```
QY 297 FGASCVTACPNYLVSTDVSGCTLVCPHNOEVTAEADGTOR-CEKSKPKCARVCYGLGMEH 355
Db 279 YGSICVSOQPTHFV-VDGSVCVCPDPDKMEV--ERGSQRCCLSGCLCPKVCCEGTGAE- 334
QY 356 LREVRVAVTSANTQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLOVFEETLEBITGY 415
Db 335 --QRQTVSSNDISFNCTKIQSLHFLVTGILGDDFKVPPDADAKKLEVFRTVREITDI 392
QY 416 LVISAMPDLPDLSVFQNLQVIRGRILHNGAVSLTLQGLGISWLGSLRSRLGSLGALIH 475
Db 393 LNIQSPKELNDLSVFSSLTIIQGRSLFKRFSLMVMRIPTLTSLGSLRSREISGVSVIS 452
QY 476 HNTILCFVHTVPWDQLFRNPH-QALLHTANRDECEGGLACHQOLCARGHCWGPPQTC 534
Db 453 QNAHLCYHHTVNWTLQFRGSRVANSLSNRMAECVADGRVCDPLCSDSGCWGPGPDQC 512
QY 535 VNCSPFLRGQECVEECRLQGLPREYVNAH-CLPCHPECOPONGSVTCFGEADOCVAC 593
Db 513 LSCRYSRHTCVAGCHFNSGIPREPAGLNGVCVACHPECKPQTGKASTGPGADCEMAC 572
QY 594 AHYKDPFPCVARCPSGVKPDLSYMPIWKFPDDEGACQPCPINCTHSCVDLDKDKCPAEQR 653
Db 573 TKFRDGPYCMSSCPAGVN-DGEKGLIFKFPNREHCEPCHQNTQCGSGPGLNDC--LE 628
QY 654 ASPLTSIVSAVVGILLVVLGVVF-----GLIKRQOKIRKYTMRLLOQETELVEPL 706
Db 629 AARLTSSQITGIALGVPAIGLFLVFLGLMGLYHRLAIRRKRAMRRYLSGESFEPL 688
QY 707 TPGAMPNQAOMRIKETELRKYVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTS 766
Db 689 GP-GEKGTVKHARILKPSDLRIKPLGSGVGVGTSGFWIPGETVKIPVAIKTQDSSG 747
QY 767 PKANKEILDEAYVMAGVGSYVSRLLGICLTSTVLQVLTQMPYGCILLDHVRENRLGSG 826
Db 748 RQTFEITDHLSSMSLDHPYVRLIGICPGTCLQVLTSSSHGSLLEHIRQHTSLDQ 807
QY 827 DLLNMCQIAKMSVLEDLVRLDLAARNVLKSNHVKITDFGLARLLDDETEYHAD 886
Db 808 RULNMCVQIAKMYLEEHVRVHKNLAARNILLKNDYQVOISDYGVDLILYDDKKYVYS 867
QY 887 CGKVPKIMMALESILRRRTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGER 946
Db 868 ETKTIKIMMALESILFRYTHQSDVMSYGVTVWEMSGFAEPYASVQPEVSVLEKGER 927
QY 947 LPQPICTIDVIMVWKMIDSECRPRELVSEFSRMDARPQRFVVIQNEDLGPASPL 1006
Db 928 LSQPAICTIDVIMVWKMIDENIRPTFKELASDFTRMARDPPRYLVIRMEG----- 980
QY 1007 DSTFYRSLLEDDMDGLVDAAEYLVPOQGFCCPDAPGAGMVHHRSSSTRSGGDLT 1066
Db 981 -----EDSGMGEFL-----RGSER---GLLE 999
QY 1067 LGLEPSEBAPRSLAPSEGAGSDVFDGLGMG---AAKGLQSLPHDPSPLQ----- 1116
Db 1000 ADLEDEEB-----GLGDRFATPSLQPSFSWSTSPSQINSYMW 1038
QY 1117 ---RYSEDPTVPLPSETGYVAPLTCSPQ- EYVQ-----PDVRPQPPSP 1158
Db 1039 TQLRYD-----FAVSQGHIGYLPWSPSPVDITIRQLWYQSRSLSSVTLPLDRSAFRSS 1092
QY 1159 REGPL- -PAARPAGATLERAKTLSPGKGVKDVFAFGGAVENPEYLPQGGAAQPPHP 1216
Db 1093 REAECEGAQACAGIFRV-----FGSERGN-----PQGG----- 1122
QY 1217 PAFSPAFLNYWDQDPPPERGAPPSTFKTPTAENPE 1253
Db 1123 -----QQRKLSTASSPSSFXTWADEEDE 1146

RESULT 7
Q9BIH9 PRELIMINARY: PRT: 1433 AA.
AC Q9BIH9;
```

```
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RP SEQUENCE FROM N.A.
RC STRAIN=SUA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.9%; Score 2047.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred. No. 2e-145;
Matches 476; Conservative 195; Mismatches 389; Indels 395; Gaps 37;

QY 26 CTGTDMLKRLRASPETHLDMLRHLVGGCQVQGNLELTLYPTVASLSFLQDIOEVQGYVL 85
Db 1 CIGTNGRMSVPANREYHYKNLRDRYTNCTYVDGNLEITWQITDNLFLQHIREVTVGYVL 60
QY 86 TAAHNOVRQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGLREL 140
Db 61 ISLYLDPQVILPRLQIIRGRTTFKLKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTEILKGGVLIQRNPQLCYQDTILWKQFNFTVSFMLRVKVSASHLENRS---RA 197
Db 105 ELPALRDILGSGVGFNNYLNCHKMSINWEEI-----LLAPQTSMQYTFNFSSPERV 156
QY 198 CHPCSPMKGSRSCWGESSEDCQSLRTVTCAGCA--RCKGPLPTDCCHCQCAAGCTGPKH 255
Db 157 CPPCHPSCEVG--CWEGEHAHQCFKJNCSPQSCQRCFGPKPRECCFLFCAGGCTGPTQ 215
QY 256 SDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLVSTDV 315
Db 216 SDCLACKNPFYDDGVCKQECPPMQIYNPTNYFEPNPDGKYAGATCVRKCP-EHLKDN 274
QY 316 SCTLVCPHNOEVTAEADGTORCEKSKPKCARVCYGLGMEHLREVRVAVTSANTQEFAGCKK 375
Db 275 ACVRKCPKGMQPONSE-----CVPCKGVCPKTCPEGIVH-----SDNIGNYKDCITI 321
QY 376 IFGSLAFLPESFDGDPASNT-----APLOPEOLOVFEETLEBITGYLVISAMPDLPD 427
Db 322 IEGSLEILDQSGDFGQQVYTNFSGPRYIKIDPDRLEVFSTVKETITGTFINIQAHHPNFTT 381
```

```
Qy 428 LSVFQNLQVIRGRILHNGAY-SLTLOGLGISWLGRLSRLRELGLGLALIHHTHLCFVHTV 486
Db 382 LNTFRNLEVGGRLKENLFASVTIVKTSLSLEKSLKRVNSGSIIVILENSDLCFVEDI 441
Qy 487 PWDLFRNPQALHTANRPEDECVCBGLACHQLCARGHCWGPGPTQCVMNCSQLRQEC 546
Db 442 DWSEIKKSSDHEVMVQKNRNATECHEGMECEQCSKAGCWGKGPEQCLECKNVKYGKC 501
Qy 547 VEBCRVLOGLPREY-VNARHCLPCHPECOPONGSVTCFGEADOCVACHYKDPFPVCAR 605
Db 502 LDSCK---SLPRLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSMNVKQGRFCVAE 553
Qy 606 CP-----SGVPRDLSYPMIPKFPD 624
Db 554 CPTTKHAWGTCINCHKTCVCGCRPRDTIAPDGCISCDAKILGSDAKIERCLMKDSCPD 613
Qy 625 -----BEG----- 627
Db 614 GYSDYVLQEBGLQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGECQECDECP 673
Qy 628 -----ACQCPINCT-----HSCVDL-----DD-----KCPAEQ- 652
Db 674 QDFYANETRCLPCHQECRGCHGLGDHHECRNLKLFEGDPYDNATFTTCVSNCPASHP 733
Qy 653 -----RASPLTSIVSAVVGILLVVVLGVVFGI---LKKRQ 685
Db 734 YKRFPEAGKIGPYCSADSMQSGRIEPTQVKIWMGSMVALIILCVVFGIAFLVFRHK 793
Qy 686 QKIRKYTMRLLOTELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAGFTVYKGIW 745
Db 794 NKDVAKMTALACGEDSEPLRPSNVGNPLTKRIKEAEIRRGVGLMGAGFVRPKGVW 853
Qy 746 IPDGNVKIIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSVLQVTO 805
Db 854 MPGESVKIIPVAIKVLMEMSGSESKFLEAYIMASVEHFNLLKLLAVCMTSQMLITQ 913
Qy 806 LMPYGCLLDHVRENRLGSDOLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHV 865
Db 914 LMPYGCLLDYVRNKKIGSKALLNWSQIARGWAYLEERLVLHRDLAARNVLVQTPSCV 973
Qy 866 KITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRPTHQSDVMSYGVTVWELMTFG 925
Db 974 KITVFLAKLLDFDSDEYRAAGGKMPKWLALCEIRHRVFTSKSDVMVAFGITIWEELTYG 1033
Qy 926 AKPDGIPAREIPDLLEGERLPPOPICTIDVYIMVVKWIMIDSECRPRELSEFSRM 985
Db 1034 ARPENYPAKDVPPELIEIGHKLQPDICSLDVCILSCWVLDADARPTFKQLAETFAEK 1093
Qy 986 ARDPQRFVVIQNEGLPASPLDSTFYRSLLEDDDDMGDLV----- 1024
Db 1094 ARDPGRYLM-----PGDKFMRLPSTYNQDEKDLIRLAPVMAAAAAAAGASN 1144
Qy 1025 -----DAEYLVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLPEPSEEA 1076
Db 1145 VDVPSITAEITDEVLPKTRPSIMLPGPSA-----VEPS-DEM 1180
Qy 1077 PRS-----PLAP---SEGAGSDVPDGLNGCAKGLQSLPTHDPSPLORYSEDPVPLP 1127
Db 1181 PKSLRYCKDLPKPDDETDHGKEV-----GVGGIR-----LNLP 1214
Qy 1128 SETDGYVAPLTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGW 1187
Db 1215 LDEDDLMP-TCQSQ---NQS-----TFG-----Y 1235
Qy 1188 KDVFAGGAVENPBYL-----TFQGGAAPOPHPPPAFSPAFDNLVYWDQDPPPERGA 1238
Db 1236 MDLIGVPASVDNPEYLMGSTQAIAGLAGSMG--PHTPP----- 1272
Qy 1239 PPSTFKGTPTAENPE 1253
Db 1273 PPNTPNGMPTHQHSQ 1287
```

```
RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF177761; AAD56009.2; -.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_L domain; 1.
DR SMART: SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 26.0%; Score 1777.5; DB 4; Length 419;
Best Local Similarity 84.2%; Pred. No. 8.1e-126;
Matches 340; Conservative 11; Mismatches 32; Indels 21; Gaps 5;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNVQVPLQRLRIVRGTFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEQVGYVLIAHNVQVPLQRLRIVRGTFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTCASPGGLRELQSLTEILKGGVLIQRNPQLCYQDTILWKDF---NNPTV 177
Db 121 DPLNNTPTVTCASPGGLRELQSLTEILKGGVLIQRNPQLCYQDTILWKDF--- 177
Qy 178 SFWLRVPKVSASHLE-NRSRACHPCSPMKGSCRCWGSESSDCQSLTRTVCAAGCARCKGP 236
Db 178 -----QLALTLLDTRSRACHPCSPMKGSCRCWGSESSDCQSLTRTVCAAGCARCKGP 230
Qy 237 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPYNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMHLL 356
Db 291 FGASCVTACPYNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCAR-----GTHSL 345
Qy 357 REVRAVTSANIOEPAGCKKIFGSLAFLPESFDGDPASNTAPLQ 400
Db 346 PPRPAVVPVPIRMQPG--PAHPVLFLRPSWDLVSIFYSLPLAP 387

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.4e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPQPPICT 954
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLEKGERLPQPPICT 60
QY 955 IDVYMTVMKWMIDSECRPRFRELVSFSEFMRMDPQRFVVIQNEIDLGPASPLDSTFYRSL 1014
Db 61 IDVYMTVMKWMIDSECRPRFRELVSFSEFMRMDPQRFVVIQNEIDLGPSSPMDSTFYRSL 120
QY 1015 LEDDDMGDLVDAEYLVPOGGFPCDPAPGACGMVHRHRSSTRSGGDLTLGLEPSEE 1074
Db 121 LEDDDMGDLVDAEYLVPOGGFPCDPALGTGTAHRRHRSARSGGELTLGLEPSEE 180
QY 1075 EAPRSLAPSEAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGYV 1134
Db 181 EAPRSLAPSEAGSDVFDGDLAVGVTKGLQSLSPHDLSPQLQRYSDPTLPLPETDGYV 240
QY 1135 APLTCSPOEYVNOQVDRPQPSREGPIPAARPAGATLERAKTLPFGKNGVVKVFAFG 1194
Db 241 APLTCSPOEYVNOQVDRPQPSREGPIPAARPAGATLERAKTLPFGKNGVVKVFAFG 300
QY 1195 GAVENPEYLTPOGGAPOHPHAPFAFDNLVYWDPPERCAPSTFKGTPTAENPEY 1254
Db 301 GAVENPEYLTPOGGAPOHPHAPFAFDNLVYWDPPERCAPSTFKGTPTAENPEY 360
QY 1255 LGLDVEV 1261
Db 361 LGLDVEV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.1e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PQNGSVTCFGEADQCACAHYKDPFCFVARCPSGVKPDLSYMPIWKFPEEGACQPCPI 634
Db 141 PEETATPKTGP--DHCWKCAHFIDGPHCVKACAGVLENDTTL-VWKYADANAVCOLCHP 197
QY 635 NCTHSCVDLDDKGCAPAEQASPLTSIVSAV--GILLVVVLGVVFGTILIKERQOKIRKVTM 693
Db 198 NCTRGCKGFGLEGCP---NGSKTPSTAAGVVGGLLCLVVGVLGILGLYLRRL-HIVRKRTL 253
QY 694 RRLQETELVEPLTPSGAMPNQAMRIKTELKRVKVLGSGAGFTVYKGIWIPDGENVK 753
Db 254 RRLQERELVEPLTPSGEAPNQAHILAKETEFKVKVVLGSGAGFTVYKGLMPEGEKVK 313
QY 754 IPVAIKVLENTSPKANKELIDBAYVMAGVSPYSRLIGLICLTSTVQLVTQMPYGCLL 813
Db 314 IPVAIKELREATSPKANKELIDBAYVMASVDNPRVCRLLIGLICLTSTVQLITQMPYGCLL 373
QY 814 DHVRENRRGLSGODLLNWCQAKGNSYLEDVLRDLAARNVLKSPNHVKITDRGLA 873
Db 374 DYIREHKDNIGSQYLLNWCQAKGNSYLEDVLRDLAARNVLKSPNHVKITDRGLA 433
QY 874 RLIDIDETEHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGP 933
Db 434 KLLGADEKHYHAGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGP 493
QY 934 AREIPDLLKGERLPQPICTIDVYIMVKWMIDSECRPRFRELVSFSEFMRMDPQRFV 993
Db 494 ASEISSVLEKGERLPQPICTIDVYIMVKWMIDSECRPRFRELVSFSEFMRMDPQRFV 553
QY 994 VIQ-NEDLGPASPLDSTFYRSLLEDGDLVDAEYLVPOGGFPCDPAPGAGGVVHR 1052
Db 554 VIQGDHMLPSPDTSKFTYRLMEEEDMEDI VDAEYLVPHOGFF----- 598
QY 1053 HRSSTRSGGDLTLGLEPSEBEAPRSL-----APSEAGSDVFDGDLGMAAGLQSL 1107
Db 599 -NSPST-----SPTPLSSLSATSNNSATNCID-----RNCQGH 631
QY 1108 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPOEYVNOQVDRPQPSREGPLPA 1165
Db 632 PVREDSFVQRYSSDPTGNFLESIDGDL-----PAPEYVNO--LMPKKPS----- 675
QY 1166 ARPAGATLERAKTLPFGKNGVVKDVF-----AFGGAVENPEYL 1203
Db 676 -----TAMVQNOIYNNISLTATSKLPMDSRYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
```

RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCACAHYKDPFCVARGSPGVKPLSYMPIWKFPEDEGACQPCPINTCHSCVDL 643
DB 1 GP--DHCMCKAHFTDGHPCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCCTRGCKGP 57

QY 644 DDGCGPQAEQASPLTSIVSAV-GILLVVLGVVFGILLIKRQOKIRKYMRLLQTEL 702
DB 58 GLEGG---NGSTPTSAAGVWGLLCLVWGLGIGLYLRR-HIVKRTLRLLQREL 113

QY 703 VEPLTPSGAMPQAOAKILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVL 762
DB 114 VEPLTPSGEAPNQAHLRLAETEFKVKVLGSGAFGVYKGLWIPGEKVKIPVAIKEL 173

QY 763 ENTSPKANKIILDEAYVMAGVGSYVSRLLGICLTSTVQLVQLMYPGCLLDVHRENR 822
DB 174 EATSPKANKIILDEAYVMASVNDPRVCRLLGICLTSTVQLITQLMYPGCLLDVIREHKN 233

QY 823 LGSQDLNMCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNMCVQIAKNVYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADKE 293

QY 883 YHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942
DB 294 YHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353

QY 943 KGERLPQPICTIDVYIMVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPICTIDVYIMVKCWMIDADSRPKRELIAEFSKWARDPPRYLVIGQDERMH 413

QY 1002 PASPLDSTFVRSLEDDMDGLVDABEYLVPOQGFCCPDPAAGAGGVHRRSSSTRSG 1061
DB 414 LPSPTDSKFRVRLMEEDMDIVDABYLVPHQGF-----NSPST--- 454

QY 1062 GGLTLGLEPSEEAAPRSL-----APSEGAGSVDFDGLCMGAAGLQSLIPTHDPSPLO 1116
DB 455 -----SRTPLLSLSLATSNNATNCID-----RNGQGHVPYEDSFVQ 491

QY 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPGATLE 1174
DB 492 RYSSDPTGNFLESIDGFL-----PAPEYVNO--LMPKKPS----- 526

QY 1175 RAKTLSPGKGVKDVFE-----AFGGAVENPEYL 1203
DB 527 -----TAMVQNOIYNNISLTAISKLPMSRYSQNSHTAVDNPYL 566

RESULT 12
Q8WYVO
ID Q8WYVO PRELIMINARY; PRT; 412 AA.
AC Q8WYVO;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN Hypothetical 44.7 kDa protein.
PF3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; EF HAND; UNKNOWN 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 8.8e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60

QY 955 IDVYIMVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQEDLGPASPLDSTFYRSL 1014
DB 61 IDVYIMVKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQEDLGPASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDABEYLVPOQGFCCPDPAAGAGGVHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDABEYLVPOQGFCCPDPAAGAGGVHRRSSSTRSGGDLTLGLEPSEE 180

QY 1075 EAPRSLAPSEGAGSVDFDGLCMGAAGLQSLIPTHDPSPQRYSEDPVPLPSETDGYV 1134
DB 181 EAPRSLAPSEGAGSVDFDGLCMGAAGLQSLIPTHDPSPQRYSEDPVPLPSETDGYV 240

QY 1135 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPGATLERAKTLSPGKGVKDVFAFG 1194
DB 241 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPGATLERAKTLSPGKGVKDVFAFG 300

QY 1195 GAVENPEYLTPOGGAAPQP-----HPPPA---FSPAFDNL 1226
DB 301 GAVENPEYLTPOGGAALSPDLLPSAQPSSTTGTTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YYWD-QDPPR-----GAPPSFKGTPTAEN 1251
DB 361 WWTQCEPEQGVRRSPDVSSSGSREGLTSAGIKRWEGPPTTSRGTTCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.

```
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruckin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24 (1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CA336459.1; -
DR EMBL; X52211; CA336459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhmrn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TYRK; 1.
DR SMART; SM00199; ZNF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.1%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.5e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 547 VEECRVLQGLPRE-VYNAR-HCLP-----CHPEQC 574
DB 354 I E K Q E S Y L L A F E I N Y R K N I P H F W S K L L M K V A D L R M I G A Y H A S R F L H M K V E C P T E L S 413
QY 575 P Q N G S V T C F G P E A D C V A C A H Y K D P P F C V A R C P S G V K P D L S Y M P I W K F P D E S G A C O P C P I 634
DB 414 P Q E -----V G P --D H C M K C A H F I D G P H C V K A C P A G V L G E N D T L -V M K Y A D A N A V C O L C H P 465
QY 635 N C T H S C V D L D D K G C P A E Q R A S P L T S I V S A V V -G I L L V V L G V V F G I L I K R R O O K I R K Y T M 693
DB 466 N C T R C K G P L E G C P ---N G S K T P S I A A G V V G L L C L V V V G I G I L Y L R R R -H I V R K R T L 521
QY 694 R L L Q E T E L V E P L T S G A M P N Q A O M R I L K E T E L R K V K L G S A F G T V Y K I W I P G E N V K 753
DB 522 R L L Q E R E L V E P L T S G E A P N Q A H L R I L K E T E F K V K V L G F G A F G T V Y K L W I P E G E K V T 581
QY 754 I P V A I K V L R E N T S P K A N K E I L D E A Y V M A G V G S P Y V S R L L G I C L T S T V Q L V T O L M P Y G C L L 813
DB 582 I P V A I K E U R A T S P K A N K E I L D E A Y V M A S V D N P H V C R L L G I C L T S T V Q L I T O L M P Y G C L L 641
QY 814 D H V R E N R G L S Q D L L N C M Q I A K G M S Y L E D V R L V H R D L A A R N V L V K S P N H V K I T D F G L A 873
DB 642 D Y I R E H K N I G S Q Y L L N C V Q I A K G M N Y L E E R H V H R D L A A R N V L V K T P Q H V K I T D F G L A 701
QY 874 R L L D I D E T Y H A D G K V P I K W A L E S I L R R R T H O S D V M S Y G V T V W E L M T F G A K P Y D G I P 933
DB 702 K Q L G A D E K Y H A E G G K V P I K W A L E S I L H R I Y T H O S D V M S Y G V T V W E L M T F G S K P Y D G I P 761
QY 934 A R E I P D L L E K E R L P Q P I C T I D V Y M I W K M I D S E C R P R E L V S F S R N A R D P Q R F V 993
DB 762 A S E I S S V L E K E R L P Q P I C T I D V Y M I W K M S G A D S R P K R F R E L I A E F S K M A R D P P R Y L 821
```

```
QY 994 V I Q -N E D L G P A S P L D S T F Y R S L L E D D M G D L V D A E Y L V P Q O G F F C P D P A P G A G M V H R 1052
DB 822 V I Q D E R M H L P S P T D S K F Y R T L M E E D M E D I V D A D E Y L V P H Q G F F ----- 866
QY 1053 H R S S T R S G G D L T L G L E P S E A P R S P L A P S E G A G S V D P D G L G M A A K G L S L P H O P 1112
DB 867 -N S P S T-----S R T P L L S L S A T S N-----N S A T K C I D R N G G H-- 898
QY 1113 S P L Q R Y S E D P T V P L S E T D G Y V A P L T C S P Q E V V N Q P D V R P O P S P R E G L P A A R P A G A T 1172
DB 899 -----P V R E D G F L-----P A P E Y V N Q--L M P K P S T A M V Q N Q I Y N Y I S L T 936
QY 1173 -L E R A K T L S P G K N G V V K D V F A F G A V E N P E Y L 1203
DB 937 A I S K L P M D S R Y Q N-----S H S T A V D N P E Y L 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=28971102;
RA Scotting P., Vennetrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278 (1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CCA0F8AF4 CRC64;

Query Match 24.0%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 G E A D C V A C A H Y K D P P F C V A R C P S G V K P D L S Y M P I W K F P D E G A C O P C P I N C T H S C V D L 643
DB 1 G P --D H C M K C A H F I D G P H C V K A C P A G V L G E N D T L -V M K Y A D A N A V C O L C H P N C T R C K G P 57
QY 644 D D K G C P A Q R A S P L T S I V S A V V -G I L L V V L G V V F G I L I K R R O O K I R K Y T M R R L L Q E T E L 702
DB 58 G L E G C P ---N G S K T P S I A A G V V G L L C L V V V G I G I L Y L R R R -H I V R K R T L R R L L Q E R E L 113
QY 703 V E P L T P S G A M P N Q A O M R I L K E T E L R K V K L G S A F G T V Y K I W I P G E N V K I P V A I K V L R 762
DB 114 V E P L T P S G E A P N Q A H L R I L K E T E F K V K V L G F G A F G T V Y K L W I P E G E K V T I P V A I K E U R 173
QY 763 E N T S P K A N K E I L D E A Y V M A G V G S P Y V S R L L G I C L T S T V Q L V T O L M P Y G C L L D H V R E N R G R 822
DB 174 E A T S P K A N K E I L D E A Y V M A S V D N P H V C R L L G I C L T S T V Q L I T O L M P Y G C L L D Y I R E H K N 233
QY 823 L G S Q D L L N C M Q I A K G M S Y L E D V R L V H R D L A A R N V L V K S P N H V K I T D F G L A R L L D I D E T E 882
DB 234 I G S Q Y L L N C V Q I A K G M N Y L E R H L V H R D L A A R N V L V K T P Q D V K I T D F G L A Q L G A D E K E 293
```


Search completed: July 22, 2003, 09:00:33
Job time : 53.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-149-163-14

Perfect score: 6848

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6715	98.1	1255	21	Human heregulin 2
2	6715	98.1	1255	22	Human tyrosine kin
3	6715	98.1	1255	22	HER2 transgene pla
4	6715	98.1	1255	23	Human HER2 (ErbB2)
5	6709	98.0	1255	17	AAW01111
6	6709	98.0	1255	20	AAW92406
7	6709	98.0	1255	21	AAW21198
8	6709	98.0	1255	21	Human HER-2/neu pr
9	6709	98.0	1255	22	Amino acid sequenc
10	6709	98.0	1255	22	Human HER-2/neu pr
					HER2/neu amino aci

11	6709	98.0	1255	23	AAE24067	Human Her-2 protei
12	6709	98.0	1255	23	AAE20479	Human Her-2/neu pr
13	6709	98.0	1255	23	AAW51143	Human Her-2/neu on
14	6709	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6666	97.3	1433	14	AAAR39568	Sequence of c-erbB
16	6545	95.6	1223	13	AAAR98923	Human breast canc
17	6392	93.3	1200	21	AAAB21208	Human HER-2/neu pr
18	5928.5	86.6	1256	21	AAAB21199	Rat Her-2/neu prot
19	5928.5	86.6	1256	23	AAW51144	Mouse Her-2/neu pr
20	5912.5	86.3	1256	21	AAAB21206	Mouse Her-2/neu pr
21	5912.5	86.3	1256	22	AAAG2860	Amino acid sequenc
22	5912.5	86.3	1256	23	AAW51151	Mouse Her-2/neu on
23	4795	70.0	919	21	AAAB21203	Human HER-2/neu fu
24	4795	70.0	919	23	AAW51148	Her-2/neu extracel
25	4062.5	59.3	920	23	AAW51152	Mouse Her-2/neu ex
26	4062.5	59.3	926	23	AAW51153	Mouse Her-2/neu ex
27	3679	53.7	712	21	AAAB21204	Human HER-2/neu fu
28	3679	53.7	712	23	AAW51149	Her-2/neu extracel
29	3533	51.6	782	18	AAW19764	Her2-GM-CSF immuno
30	3531	51.6	653	21	AAAB21200	Extracellular HER-
31	3531	51.6	653	23	AAW51145	Human Her-2/neu on
32	3493	51.0	645	22	AAAB60408	Human ErbB2 oncopr
33	3493	51.0	645	22	AAAB61593	Human ErbB2 extrac
34	3428	50.1	951	21	AAV44993	DC8ecFu-erbB2EC fu
35	3325	48.6	624	11	AAAR08222	Extracellular port
36	3164	46.2	1210	21	AAAB19259	Amino acid sequenc
37	3164	46.2	1210	21	AAAY0616	Human EGF receptor
38	3164	46.2	1210	23	AAE23019	Human Her-1 protei
39	3164	46.2	1210	23	AAW50768	Human epidermal gr
40	3162	45.2	1210	22	AAAB68420	Amino acid sequenc
41	3123	45.6	1210	23	ABP51768	Human epidermal gr
42	3084	45.0	583	23	AAE20483	Human protein for
43	3084	45.0	587	23	AAE20481	Human protein for
44	3083	45.0	589	23	AAE20484	Human protein for
45	3083	45.0	600	23	AAE20482	Human protein for

ALIGNMENTS

RESULT 1

AA92620
ID AA92620 standard; Protein; 1255 AA.

XX
AC AA92620;

XX
DT 10-AUG-2000 (first entry)

XX
DE Human heregulin 2 (Her2).

XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;

KW self-protein; cancer; breast cancer; prostate cancer;

KW cell-associated peptide antigen; foreign epitope.

OS Homo sapiens.

XX
FH Key Location/Qualifiers

FT Domain 1..173 /label= N-terminal

FT Region 5..25 /note= "mature polypeptide"

FT Region /label= insertion region /note= "suitable for foreign epitope insertion"

FT Region 59..73 /label= insertion region /note= "suitable for foreign epitope insertion"

FT Region 103..117 /label= insertion region /note= "suitable for foreign epitope insertion"

FT Region 149..163 /label= insertion region /note= "suitable for foreign epitope insertion"

FT Domain 174..323 /label= insertion region /note= "suitable for foreign epitope insertion"

Qy 777 AYVWAGVSPVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRRGLSQDQLNNWCMQIA 836
Db 771 AYVWAGVSPVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRRGLSQDQLNNWCMQIA 830
Qy 837 KGMSVLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMA 896
Db 831 KGMSVLEDRVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMA 890
Qy 897 LESILRRRFTHQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICITID 956
Db 891 LESILRRRFTHQSDVWSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGERLPQPPICITID 950
Qy 957 YVIMVWKCWIMIDSECRPRFRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 YVIMVWKCWIMIDSECRPRFRELVSFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDWMDLVDAEYLVPOQGFPCDPAPGAGGVVHHRSSSTRSGGDLTLGLEPSEBEA 1076
Db 1011 DDDWMDLVDAEYLVPOQGFPCDPAPGAGGVVHHRSSSTRSGGDLTLGLEPSEBEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPAFGGA 1196
Db 1131 LTCSPQPEYVNPQVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPAFGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note= "Antigenic epitope"
XX
PN WO200168677-A2.
XX
PD 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
PR 16-MAR-2000; 2000US-0527487.
XX
PA (GENZ) GENZYME CORP.
XX
PI Nicolette CA;
XX
DR WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX

PT Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT .
XX Claim 4; Page 63-67; 69pp; English.
PS
XX The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6715; DB 22; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;
Qy 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPETHLDMRLHYQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPETHLDMRLHYQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVFLRLRI VRGTQLFEDNALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVFLRLRI VRGTQLFEDNALAVLDNG 120
Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCVQDTILWKDF---NNFTV 177
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCVQDTILWKDF---NNFTV 177
Qy 178 SFWLRVPKVSASHLE-NRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGCARCKGP 236
Db 178 -----QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGCARCKGP 230
Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESHPNPGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESHPNPGRYT 290
Qy 297 FGASCVTACPNYLSLTDVGSCTLVCPLNQVTAEDGTORCEKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLSLTDVGSCTLVCPLNQVTAEDGTORCEKSKPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANTIOEPAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFTELEETGYL 416
Db 351 REVRAVTSANTIOEPAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFTELEETGYL 410
Qy 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHH 476
Db 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRELGSGLAIHH 470
Qy 477 NTHLCFVHTVPWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVN 536
Db 471 NTHLCFVHTVPWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVN 530
Qy 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHY 596
Db 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHY 590
Qy 597 KDPFFCVARCPGSKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPORASP 656
Db 591 KDPFFCVARCPGSKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPORASP 650

```
QY 657 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 716
DB 651 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 776
DB 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVQLVLTQMLPVGCLLDHVRENRRGLGSDLLNWCQOIA 836
DB 771 AYVMAGVGSPPYVSRLLGICLTSTVQLVLTQMLPVGCLLDHVRENRRGLGSDLLNWCQOIA 830
QY 837 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWWA 896
DB 831 KGMYSILEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWWA 890
QY 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
DB 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
QY 957 VYIMVUKWIMIDSECRPRFRELSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLE 1016
DB 951 VYIMVUKWIMIDSECRPRFRELSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLLE 1010
QY 1017 DDDMGDLVDAEYLVLPQGGFFCDDPAPGAGGMVHHRHSSTRSGGGDLTLGLEPSEEEA 1076
DB 1011 DDDMGDLVDAEYLVLPQGGFFCDDPAPGAGGMVHHRHSSTRSGGGDLTLGLEPSEEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVYAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVYAP 1130
QY 1137 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGA 1196
DB 1131 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGA 1190
QY 1197 VENPEYLTPOGGAAPHPHPPAPSPAFDNLYYWDQDPPPERGAPPTFKGTPTTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPHPHPPAPSPAFDNLYYWDQDPPPERGAPPTFKGTPTTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC
XX
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
XX Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
XX 04-JAN-2001.
PD
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson S, Schwall R;
PI
```

```
XX WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid.
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;

Query Match 98.1%; Score 6715; DB 22; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;

QY 1 MELAALCRNGLLIALLPGAASSTQVCTGDMKRLRASPETHLDMLRHLHYGGCVVQGNL 60
DB 1 MELAALCRNGLLIALLPGAASSTQVCTGDMKRLRASPETHLDMLRHLHYGGCVVQGNL 60
QY 61 EUTYLPNLSLFLQDIEVQGVYLIHNVQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
DB 61 EUTYLPNLSLFLQDIEVQGVYLIHNVQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177
DB 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177
QY 178 SFMLRPVKVSASHLE-NRSRACHPCSPCKSGSCWGESSEDCOSLTRTVCAGCARCKGP 236
DB 178 -----QLALTIDTNRSRACHPCSPCKSGSCWGESSEDCOSLTRTVCAGCARCKGP 230
QY 237 LPTDCCHEOCAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
DB 231 LPTDCCHEOCAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYVLTSDVGSCTLVCPHNOEVTAEQDQCEKSCPKARVCVGLGMEHL 356
DB 291 FGASCVTACPNYVLTSDVGSCTLVCPHNOEVTAEQDQCEKSCPKARVCVGLGMEHL 350
QY 357 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 416
DB 351 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGSLRLSLGSLALIIHH 476
DB 411 YISAWPDSLPLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGSLRLSLGSLALIIHH 470
QY 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCGWGPCTOCVN 536
DB 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCGWGPCTOCVN 530
QY 537 CSQFLRGQCEVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 596
DB 531 CSQFLRGQCEVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 590
QY 597 KDPFFCVARCPGVKPDLSYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGCBAEQRAS 656
DB 591 KDPFFCVARCPGVKPDLSYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGCBAEQRAS 650
QY 657 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 716
DB 651 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYIMRRLLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 776
DB 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKBILDE 770
```

```
Db      711 QMRILKTELKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDE 770
QY      777 AYVMAGVSPYVSRLIGICLTSTVOLVTQLMPYGCLLDHVRENRLGSDLLNWCQIA 836
Db      771 AYVMAGVSPYVSRLIGICLTSTVOLVTQLMPYGCLLDHVRENRLGSDLLNWCQIA 830
QY      837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 896
Db      831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 890
QY      897 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db      891 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
QY      957 VYIMVYKWMIDSECPRELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016
Db      951 VYIMVYKWMIDSECPRELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010
QY      1017 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEA 1076
Db      1011 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLEPSEEEA 1070
QY      1077 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db      1071 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY      1137 LTCSPQEVYNQDVRPQPSREGPLPAARPAATLERAKTLSPGKNGVVKDVPFPGA 1196
Db      1131 LTCSPQEVYNQDVRPQPSREGPLPAARPAATLERAKTLSPGKNGVVKDVPFPGA 1190
QY      1197 VENPEYLTTPQGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLG 1256
Db      1191 VENPEYLTTPQGAAPQHPHPPAPSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLG 1250
QY      1257 LDVPV 1261
Db      1251 LDVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
```

```
PI Erickson S, Schwall R, Sliwkowski M;
XX
DR WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6715; DB 23; Length 1255;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1242; Conservative 6; Mismatches 3; Indels 14; Gaps 3;
QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTMKRLRASPETHLDMLRHLHYGQGVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTMKRLRASPETHLDMLRHLHYGQGVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIEQVGVLIYAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGVLIYAHNQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDF---NNFTV 177
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDF---NNFTV 177
QY 178 SFMLRVKPVYSASHLE-NRSRACHPCSPMCKSRGCMWSESSEDCQSLRTTVCAGGCARCKGP 236
Db 178 -----QLALTLDITNRSRACHPCSPMCKSRGCMWSESSEDCQSLRTTVCAGGCARCKGP 230
QY 237 LPTDCHEQCAAGCTGPKHSDCLACLUHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCLACLUHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYVLTSDVGSCTLVCPHMQVTAEDGTORCEKCKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVLTSDVGSCTLVCPHMQVTAEDGTORCEKCKSPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEEITGYL 416
Db 351 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEEITGYL 410
QY 417 YISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRELGSGLALIIH 476
Db 411 YISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRELGSGLALIIH 470
QY 477 NTHLCFVHTVPWDQLFRNPQALLHTANRPEDECYGEGLACHOLCARGHGWGPGTCVN 536
Db 471 NTHLCFVHTVPWDQLFRNPQALLHTANRPEDECYGEGLACHOLCARGHGWGPGTCVN 530
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHY 596
Db 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHY 590
QY 597 KDPPECVARCPSGVKPDLSYMPFIWKPDEBEGACPCPINCTHSCVDLDDKGCPEQRASP 656
```

Db	591	KDPFFCVARCPGKPDLSYMPWKPFDEBAGQPCPINCTHSCVDLDDKCPAEQRASP	650
Qy	657	LTSIVSAVVGILLVWLVGVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA	716
Db	651	LTSIVSAVVGILLVWLVGVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA	710
Qy	717	QMRILKETELURKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKETLDE	776
Db	711	QMRILKETELURKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKETLDE	770
Qy	777	AYVMAGVSPYVRLGICITSTVQLVTQMLPXYGCLLDHVRENRGRIGSODLLNWCNQIA	836
Db	771	AYVMAGVSPYVRLGICITSTVQLVTQMLPXYGCLLDHVRENRGRIGSODLLNWCNQIA	830
Qy	837	KGMSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMA	896
Db	831	KGMSYLEDVLRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMA	890
Qy	897	LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	956
Db	891	LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	950
Qy	957	VYIMVWKWIMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE	1016
Db	951	VYIMVWKWIMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE	1010
Qy	1017	DDMGDLVDAEYLVPOQGFCDPAPGAGVMVHRRSSSTRSGGGDLTLGLEPSBEEA	1076
Db	1011	DDMGDLVDAEYLVPOQGFCDPAPGAGVMVHRRSSSTRSGGGDLTLGLEPSBEEA	1070
Qy	1077	PRSPAPSEGAGSDVDFDGLGMAAGLQSLPTHDPSPLOYSYSEDPTVPLPSETDGYVAP	1136
Db	1071	PRSPAPSEGAGSDVDFDGLGMAAGLQSLPTHDPSPLOYSYSEDPTVPLPSETDGYVAP	1130
Qy	1137	LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA	1196
Db	1131	LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA	1190
Qy	1197	VENPEYLTPOGGAAPHPHPPAPSPAFDNLVYNDODPPERCAPPSFKGTPTAENPYLG	1256
Db	1191	VENPEYLTPOGGAAPHPHPPAPSPAFDNLVYNDODPPERCAPPSFKGTPTAENPYLG	1250
Qy	1257	LDVPV 1261	
Db	1251	LDVPV 1255	
RESULT 5			
ID	AAW01111 standard; Protein; 1255 AA.		
XX	AAW01111;		
AC	AAW01111;		
XX	01-JAN-1997 (first entry)		
DT	HER-2/neu protein.		
DE	HER-2/neu protein.		
XX	HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;		
KW	breast cancer; ovary cancer; colon cancer; lung cancer;		
KW	prostate cancer; immunisation; tumour; vaccine; vector.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	Homo sapiens.		
PH	Key Location/Qualifiers		
FT	Domain 676..1255		
FT	/label= Intracellular domain		
FT	/note= "claimed domain, useful for immunisation"		
XX	WO9630514-A1.		
PN	WO9630514-A1.		
XX	03-OCT-1996.		
PD	03-OCT-1996.		
XX	03-OCT-1996.		

PF	28-MAR-1996;	96WO-US01689.	
XX			
PR	31-MAR-1995;	95US-0414417.	
XX			
PA	(UNIW) UNIV WASHINGTON.		
XX			
PI	Cheever MA, Disis ML;		
XX			
DR	WPI; 1996-455361/45.		
DR	N-PSDB; AAT40739.		
XX			
PT	DNA encoding HER-2-neu poly:peptide(s) - used for prevention or		
PT	treatment of malignancies with which the HER-2/neu oncogene is		
PT	associated		
XX			
PS	Claim 2; Page 56-61; 7lpp; English.		
XX			
CC	Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is		
CC	the product of the HER-2/neu oncogene (see also AAT40739). The		
CC	protein is over-expressed in various cancers, including breast,		
CC	ovarian, colon, lung and prostate. The intracellular domain of the		
CC	protein can be used to immunise an animal against a malignancy with		
CC	which the oncogene is associated. The polypeptide can be produced		
CC	in transformed host cells for use in immunisation. Alternatively,		
CC	animal cells are transfected in vivo or ex vivo with a viral vector		
CC	that directs expression of the polypeptide.		
XX			
SQ	Sequence 1255 AA;		
	Query Match 98.0%; Score 6709; DB 17; Length 1255;		
	Best Local Similarity 98.0%; Pred. No. 0;		
	Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;		
Qy	1	MELAAALCRWGLLLALLPPGAASQVCTGTDKMLRSPASPETHLDMLRHLVGGCQVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASQVCTGTDKMLRSPASPETHLDMLRHLVGGCQVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIEQVGVLIHNOVQVPLQRLRIVRGTOIFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLQDIEQVGVLIHNOVQVPLQRLRIVRGTOIFEDNYALAVLDNG	120
Qy	121	DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDF---NNFTV	177
Db	121	DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDFHKN---177	
Qy	178	SFWLRVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCA	236
Db	178	-----QUALITDNRGRACHPCSPMCKSGRCWGESSEDCQSLTRTVCA	230
Qy	237	LPTDCCHCCOACAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPN	296
Db	231	LPTDCCHCCOACAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPN	290
Qy	297	FGASCVTACPNYLTSDVSGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHL	356
Db	291	FGASCVTACPNYLTSDVSGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHL	350
Qy	357	REVRVTSANIQBFAGCKIFGSLAFIPESFDGDPASNTAPLQEQVFTLSEITGYL	416
Db	351	REVRVTSANIQBFAGCKIFGSLAFIPESFDGDPASNTAPLQEQVFTLSEITGYL	410
Qy	417	YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTQGLGTSWGLRSLRELGSGLAIHH	476
Db	411	YISAWPDSLPLDSVFONLQVIRGRILHNGAYSILTQGLGTSWGLRSLRELGSGLAIHH	470
Qy	477	NTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGEGLACHQICARGHGWGP	536
Db	471	NTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGEGLACHQICARGHGWGP	530
Qy	537	CSQFLRGQECVBCRVLQGLPREYVNRHCLPCHPECPQNGSVTCFPGSADQCA	596
Db	531	CSQFLRGQECVBCRVLQGLPREYVNRHCLPCHPECPQNGSVTCFPGSADQCA	590


```
QY 597 KDPFFCVARCPGKVPDLSYMPIMKFFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQAS 656
DB 591 KDPFFCVARCPGKVPDLSYMPIMKFFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQAS 650
QY 657 LTSIVSAVGLLVVGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQA 716
DB 651 LTSIISAVGTLVVVGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDE 776
DB 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDE 770
QY 777 AYVMAVGSPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGRLGSODLLNWCQIA 836
DB 771 AYVMAVGSPYVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRRGRLGSODLLNWCQIA 830
QY 837 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDTEYHADGGKVPKXMA 896
DB 831 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDTEYHADGGKVPKXMA 890
QY 897 LESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
DB 891 LESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
QY 957 VYIMIVKCMWIDSECRPRFRELSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016
DB 951 VYIMIVKCMWIDSECRPRFRELSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010
QY 1017 DDDMGDLVDAEYLVPOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1070
QY 1077 PRSLPASEGAGSDVFDGLGMGAAGKIQSLTHDPQLQYSDPTVPLPSETDGYVAP 1136
DB 1071 PRSLPASEGAGSDVFDGLGMGAAGKIQSLTHDPQLQYSDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDYFAFGA 1196
DB 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDYFAFGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPADNLYWDQDPPERGAAPSTFKGTPTAENPEYL 1256
DB 1191 VENPEYLTPOGGAAPQHPHPPAFSPADNLYWDQDPPERGAAPSTFKGTPTAENPEYL 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255
```

RESULT 7

AAB21198
ID AAB21198 standard; protein; 1255 AA.

XX AC AAB21198;

XX DT 12-JAN-2001 (first entry)

XX DE Human HER-2/new protein.

XX KW Human; HER-2/new; oncogene; tyrosine kinase; cytostatic; vaccine;
breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX KW colon cancer.

XX OS Homo sapiens.

XX PN WO200044899-A1.

XX PD 03-AUG-2000.

XX PF 28-JAN-2000; 2000MO-US02164.

XX PR 29-JAN-1999; 99US-0117976.

XX

```
PA (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.  
PI Cheever MA, Gheysen D;  
XX WPI; 2000-505976/45.  
DR N-PSDB; AAA89736.  
XX HER-2/new extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX Claim 52; Fig 7; 128pp; English.  
XX The present sequence is the human HER-2/new protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/new  
CC gene is an oncogene. An HER-2/new fusion protein comprising a  
CC HER-2/new extracellular domain fused to a HER-2/new phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/new protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.  
XX  
SQ Sequence 1255 AA;
```

```
Query Match 99.0%; Score 6709; DB 21; Length 1255;  
Best Local Similarity 98.0%; Pred. No. 0;  
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;  
QY 1 MELAALCRWGLLIALLPGRGAASQVCTGTDKMLRSPETHLDMLRHLGYCCVVOGNL 60  
DB 1 MELAALCRWGLLIALLPGRGAASQVCTGTDKMLRSPETHLDMLRHLGYCCVVOGNL 60  
QY 61 ELTYLPTNASLSFLQDIEQVGYVLIAHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120  
DB 61 ELTYLPTNASLSFLQDIEQVGYVLIAHNOVQVPLQRLRIVRGTLQFEDNVALAVLNG 120  
QY 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQRPQLCYQDTILWKDF- --NNFTV 177  
DB 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQRPQLCYQDTILWKDF- --NNFTV 177  
QY 178 SFWLRVPKVSASHLE- NRSRACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGCARCKGP 236  
DB 178 - - - - - QLATLTIDTNRACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGCARCKGP 230  
QY 237 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296  
DB 231 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290  
QY 297 FGASCVTACPNYLTSDVGSCTLVCPHNOEVTADGTQRCCKSKPCARVCYGLGMEHL 356  
DB 291 FGASCVTACPNYLTSDVGSCTLVCPHNOEVTADGTQRCCKSKPCARVCYGLGMEHL 350  
QY 357 REYRAVTSANIQEPAGCKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEBITGYL 416  
DB 351 REYRAVTSANIQEPAGCKIFGSLAFIPESFDGDPASNTAPLOPELOVFTLEBITGYL 410  
QY 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSURELGSGLALH 476  
DB 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRSURELGSGLALH 470  
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPQTQCN 536  
DB 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPQTQCN 530  
QY 537 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 596  
DB 531 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 590  
QY 597 KDPFFCVARCPGKVPDLSYMPIMKFFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQAS 656
```

Db 591 KDPFFCVCARCGSKVDPDLSTYMPKWFDPDEGACQPCINCTHSCVDLDDKGCFAEQRASP 650
QY 657 LTSISAVVGIILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOA 716
Db 651 LTSISAVVGIILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGTYYKGIWIPGENVKIIPVAIKVLRENTSPKANKSILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTYYKGIWIPGENVKIIPVAIKVLRENTSPKANKSILDE 770
QY 777 AYVMAGVGSPPVSRLLIGLICLTSTVOLATOLMPYGCILLDHVRENRGLSQDILLNWCQIA 836
Db 771 AYVMAGVGSPPVSRLLIGLICLTSTVOLATOLMPYGCILLDHVRENRGLSQDILLNWCQIA 830
QY 837 KGMYSILEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDDIDEYHADGGKVPKMA 896
Db 831 KGMYSILEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDDIDEYHADGGKVPKMA 890
QY 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 956
Db 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 950
QY 957 VYMIWVKWIMIDSECRPRELVSFERNARDPQRFVWIONEDLGASPLDSTFYRSILLE 1016
Db 951 VYMIWVKWIMIDSECRPRELVSFERNARDPQRFVWIONEDLGASPLDSTFYRSILLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSEEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSEEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVWVDVAFGGA 1196
Db 1131 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVWVDVAFGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 8

AA484780
ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.

OS Homo sapiens.

PN WO200020579-A1.

XX 13-APR-2000.

PD 01-OCT-1999; 99WO-CA00912.

PF 02-OCT-1998; 98US-0165192.

PR 02-OCT-1998; 98US-0165192.

XX

(UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI: 2000-303768/26.

DR N-PSDB; AAA14812.

XX

PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE

PT erbb-2, inhibitors of the protein are useful for treatment of cancer

XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The

CC protein has an in-frame deletion of 16 amino acids, 2 of which are

CC conserved cysteine residues, compared to the unspliced protein. The

CC erbb-2 polynucleotide is used to construct probes for detecting

CC disorders of cell transformation such as cancer. Antibodies to the

CC protein may be used to detect SPLICE erbb-2 in a sample. Agents

CC (e.g. antisense oligonucleotides) which inhibit the expression of

CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and

CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful

CC for treating conditions involving damaged cells including conditions

CC in which degeneration of tissue occurs, such as arthropathy, bone

CC resorption, inflammatory diseases, degenerative disorders of the

CC central nervous system and wound healing.

XX

SQ Sequence 1255 AA;

Query Match 98.0%; Score 6709; DB 21; Length 1255;

Best Local Similarity 98.0%; Pred. No. 0;

Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMLRHLHLYGQGVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGDMKRLRLPASPEHLDMLRHLHLYGQGVVQGNL 60

QY 61 EUTYLPNTASLSFLQDIQEVQVYLIAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 EUTYLPNTASLSFLQDIQEVQVYLIAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

QY 121 DFLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177

Db 121 DFLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDF---NNFTV 177

QY 178 SFWLRVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGP 236

Db 178 SFWLRVPKVSASHLE-NRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGGCARCKGP 230

QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296

Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290

QY 297 FGASCVTACPNYVLTSDVGSCTLVCPHNOVTAEDGTQCEKCKPCARVCYGLGMEHL 356

Db 291 FGASCVTACPNYVLTSDVGSCTLVCPHNOVTAEDGTQCEKCKPCARVCYGLGMEHL 350

QY 357 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVETLEEITGYL 416

Db 351 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVETLEEITGYL 410

QY 417 YISAWPDSLPLDSLVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 476

Db 411 YISAWPDSLPLDSLVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIH 470

QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWGPGTQCVN 536

Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWGPGTQCVN 530

QY 537 CSQFLRGQCEVBEERVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 596

Db 531 CSQFLRGQCEVBEERVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 590

QY 597 KDPPFCVACRCPGSKVPDLSYMPDWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQRASP 656

Db 591 KDPFPCVARGSGVKPDLSTYMPIWKFPDEGACQCPINCTHSCVDLDDKCPAEQRASP 650
Qy 657 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIISAVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDE 770
Qy 777 AYVMAGVSPVSRLLGICLTSTVQLTQMLPYGCLLDHVRENRGRIGSQDLLNWCQIA 836
Db 771 AYVMAGVSPVSRLLGICLTSTVQLTQMLPYGCLLDHVRENRGRIGSQDLLNWCQIA 830
Qy 837 KGM5YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMA 896
Db 831 KGM5YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICIID 950
Qy 957 VYMIWVKCMIIDSCRRFRRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWVKCMIIDSCRRFRRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDABEYLVPQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1076
Db 1011 DDDMGDLVDABEYLVPQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLSPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLTPHDPSPLQRYSDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLTPHDPSPLQRYSDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVPFAFGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVPFAFGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
AC AAB85458;
XX AAB85458;
DT 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.
XX

PI Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
DR N-PSDB; AAH23392.
XX
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
FT for treating or preventing cancer, e.g. breast cancer
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6709; DB 22; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMLRLHYGQCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMLRLHYGQCQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNVQVQVLPQRLRIVRGTLQFEDNVVALVDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNVQVQVLPQRLRIVRGTLQFEDNVVALVDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177
Qy 178 SFMLRVPKVSASHLE-NRSRACHPCSPCKSGRSCWGESSEDCQSILTRVCAGGCARCKGP 236
Db 178 -----QLALTLDITNRSRACHPCSPCKSGRSCWGESSEDCQSILTRVCAGGCARCKGP 230
Qy 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVFTLEITGYL 416
Db 351 REVRAVTSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVFTLEITGYL 410
Qy 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGSLRSLGSLALHH 476
Db 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLSLTQGLGISWLGSLRSLGSLALHH 470
Qy 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCKGPGPTQCVN 536
Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCKGPGPTQCVN 530
Qy 537 CSQFLRQGECEVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 596
Db 531 CSQFLRQGECEVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHY 590
Qy 597 KDPFPCVARGSGVKPDLSTYMPIWKFPDEGACQCPINCTHSCVDLDDKCPAEQRASP 656
Db 591 KDPFPCVARGSGVKPDLSTYMPIWKFPDEGACQCPINCTHSCVDLDDKCPAEQRASP 650
Qy 657 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQA 716

Db 471 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVN 530
Qy 537 CSQFLRGQECVECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFPGPEADOCVACHY 596
Db 531 CSQFLRGQECVECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFPGPEADOCVACHY 590
Qy 597 KDPFFCVARCPGSKPDLSPYMPWKFPDEBAGACQPCPINCTHSCVDLDDKGCPEAORASP 656
Db 591 KDPFFCVARCPGSKPDLSPYMPWKFPDEBAGACQPCPINCTHSCVDLDDKGCPEAORASP 650
Qy 657 LTSIVANVGIILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVELPELTPSGAMPNQA 716
Db 651 LTSIIISAVVGIILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVELPELTPSGAMPNQA 710
Qy 717 QMRILKETELKURKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNENTSPKANKEITLDE 776
Db 711 QMRILKETELKURKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNENTSPKANKEITLDE 770
Qy 777 AYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGDOLLNWCQIA 836
Db 771 AYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRRGLSGDOLLNWCQIA 830
Qy 837 KGSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKGKVPKMA 896
Db 831 KGSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKGKVPKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYIMVWKWIMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKWIMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDABEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDABEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKDVFAFGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPBYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPBYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

XX AAE24067;

XX 23-SEP-2002 (first entry)

XX Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisease;
KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

PN WO200222636-A1.

XX

PD 21-MAR-2002 .
XX
XX 12-SEP-2001; 2001WO-US28572.
XX
XX 15-SEP-2000; 2000US-0663834.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Bennett CF, Cowseert LM;
XX
XX WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
XX Example 13; Page 95-107; 116pp; English.
PS
XX
XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6709; DB 23; Length 1255;

Best Local Similarity 98.0%; Pred. No. 0;

Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCQVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVGGCQVQGNL 60

Qy 61 ELTYLPTNASLSFLDIOEQVGVYVIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLDIOEQVGVYVIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILWKDF---NNFTV 177

Qy 178 SFWLVRVPKVBASHLE-NRSRACHPCSPMCKSGRCWGESSEDCSLTRTVCAAGCCKGK 236

Db 178 SFWLVRVPKVBASHLE-NRSRACHPCSPMCKSGRCWGESSEDCSLTRTVCAAGCCKGK 230

Qy 237 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTTFFSMPNPEGRYT 296

Db 231 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTTFFSMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLSLTDVSGCTLVCPHNOEVTAEADGTCRCKSKPCARVCYGLGMEHL 356

Db 291 FGASCVTACPNYVLSLTDVSGCTLVCPHNOEVTAEADGTCRCKSKPCARVCYGLGMEHL 350

Qy 357 REVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQQLQVFTLEITGYL 416

Db 351 REVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQQLQVFTLEITGYL 410

Qy 417 YISAWPDSLPDLVFNQNLQVIRGRILHNGAYSITLQGLIGISWGLRSLRELGLALIH 476

Db 411 YISAWPDSLPDLVFNQNLQVIRGRILHNGAYSITLQGLIGISWGLRSLRELGLALIH 470

Qy 477 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVN 536

Db 471 NTHLCFVHTVPMDLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCVN 530

Qy 537 CSQFLRGQECVECRVLQGLPREYVVARHCLCPCHPECOPOGNSVTCFPGPEADOCVACHY 596

```

Db 531 CSQFLRQECVEECRVLOGLPREYVNAHCLUPCHPECOQPNQSVTCFGEADQCACAHY 590
Qy 597 KDPFFCVARCFSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 656
Db 591 KDPFFCVARCFSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS 650
Qy 657 LTSIVSAVWGILLVVGWVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIISAVWGILLVVGWVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDE 776
Db 711 QMRILKTELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDE 770
Qy 777 AYVMAGVSPYVSRLLGLCLTSTVQLVTQLMPYGLLDHVRNRRGLGSDQLLNWCQIA 836
Db 771 AYVMAGVSPYVSRLLGLCLTSTVQLVTQLMPYGLLDHVRNRRGLGSDQLLNWCQIA 830
Qy 837 KGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVIKMA 896
Db 831 KGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVIKMA 890
Qy 897 LESILRRRFTHQSDVWSVGTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
Db 891 LESILRRRFTHQSDVWSVGTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
Qy 957 VYIMVWKCMIDSECRPRFRELVSFSEFMRADPQRFVVIQNEGLPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKCMIDSECRPRFRELVSFSEFMRADPQRFVVIQNEGLPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDAEYLVPQGFCDPDPAGAGGMVHRHRSSTRSGGDLTLGLLEPSEEA 1076
Db 1011 DDMGDLVDAEYLVPQGFCDPDPAGAGGMVHRHRSSTRSGGDLTLGLLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOYSDDTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPTHDPSPLOYSDDTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFARFGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFARFGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
XX
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX

```

```

PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
DR WPI; 2002-280758/32.
XX
DR N-PSDB; AAD32743.
XX
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX prevention and diagnosis of cancer, preferably breast cancer -
XX
PS Disclosure; Page 114-117; 129pp; English.
XX
CC The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

```

```

Query Match 98.0%; Score 6709; DB 23; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLKRLPASPETHLDMRLHLYGQCQVQGNL 60

Qy 61 ELTYLPTNASLSFLDIOEQVGYVLIAHNQVRQVPLORLRIVRGTQLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLDIOEQVGYVLIAHNQVRQVPLORLRIVRGTQLFEDNYALAVLNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDF---NNFTV 177
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDF---NNFTV 177

Qy 178 SFWLVRPKVSAHLE-NRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 236
Db 178 -----QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYLTSDVSGCTLVCPLNHOEVTAEQDTCRCKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLTSDVSGCTLVCPLNHOEVTAEQDTCRCKSKPCARVCYGLGMEHL 350

Qy 357 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVETEEITGYL 416
Db 357 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVETEEITGYL 416

```

351	REVRAVTSANIQEFAGCKKIFGSLAFIPESFGDGPASNTAPLQPEQLQVFPETLBEITGYL	410
417	YISAWPDSLPLDSVFONLQVIRGRIHLNNGAYSLTLQGLIGISWLGRLSRLRELGSLALIH	476
411	YISAWPDSLPLDSVFONLQVIRGRIHLNNGAYSLTLQGLIGISWLGRLSRLRELGSLALIH	470
477	NTHLCFVHTVPWDLFRNPQALLHTANRBEDECVGEGLAHQIJCARGHCWGPGPTQCVN	536
471	NTHLCFVHTVPWDLFRNPQALLHTANRBEDECVGEGLAHQIJCARGHCWGPGPTQCVN	530
537	CSQFLRGQECVEECRVLQGLPREVYNARHCLPCHPECPQONGSVTCFGRPDAQCVACAHY	596
531	CSQFLRGQECVEECRVLQGLPREVYNARHCLPCHPECPQONGSVTCFGRPDAQCVACAHY	590
597	KDPFFCVARCPSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS	656
591	KDPFFCVARCPSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS	650
657	LTSIVSAVVGILLVVVLGVVFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQA	716
651	LTSIVSAVVGILLVVVLGVVFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQA	710
717	QMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDE	776
711	QMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDE	770
777	AYNMAGVSPVSRLLGILCTSTVOLVTQLMPYGLLDHVRENRGRGLSQDLLNWCQIA	836
771	AYNMAGVSPVSRLLGILCTSTVOLVTQLMPYGLLDHVRENRGRGLSQDLLNWCQIA	830
837	KGMSYLEDLVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMA	896
831	KGMSYLEDLVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMA	890
897	LESILRRRPTHQSDVNSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLUPQPICTID	956
891	LESILRRRPTHQSDVNSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLUPQPICTID	950
957	YVIMVWKCMWIDSECRPRFRELVSFSEFMRARDPORFVVIQNEDLGPASPLDSTFYRSLLE	1016
951	YVIMVWKCMWIDSECRPRFRELVSFSEFMRARDPORFVVIQNEDLGPASPLDSTFYRSLLE	1010
1017	DDMGDLVDABEYLVPQQGFCCPDPAAGAGMVHHRSSSTRSGGGDLTLGLEPSEEEA	1076
1011	DDMGDLVDABEYLVPQQGFCCPDPAAGAGMVHHRSSSTRSGGGDLTLGLEPSEEEA	1070
1077	PRSPLASSEGAGSDVFDGLGKGAAGKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGVYAP	1136
1071	PRSPLASSEGAGSDVFDGLGKGAAGKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGVYAP	1130
1137	LTCSPQBEYVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA	1196
1131	LTCSPQBEYVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGA	1190
1197	VENPEYLYTPQGAAAPQHPPHPAFSPAFONLYYWDQDPPERCAPSTFKGTPTAENPEYLG	1256
1191	VENPEYLYTPQGAAAPQHPPHPAFSPAFONLYYWDQDPPERCAPSTFKGTPTAENPEYLG	1250
1257	LDVVPV 1261	
1251	LDVVPV 1255	

RESULT 13

AAM51143

AA51143
ID AAM51143 standard; Protein; 1255 AA.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

AC AAM51143;

DT 17-JUN-2002 (first entry)

XX
DE
XX

XX

Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185; tyrosine kinase; receptor; c-erbB2; gene therapy.

Homo sapiens.

Key	Location/Qualifiers
Domain	1..653
	/note= "extracellular domain"
Domain	676..1255
	/note= "intracellular domain"
Domain	990..1255
	/note= "phosphorylation domain"

WO200212341-A2.

14-FEB-2002.

03-AUG-2001: 2001WO-IIS24283

03-AUG-2000: 2000US-0632507

(CORI-) CORIXA CORP.
(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Cheever MA, Gheyssen D;

WPI: 2002-241743/29.

WFI; ZUZ=Z4I743/
N-PSDB: ABA92250.

Her-2/neu fusion protein for treating or preventing cancer by eliciting or enhancing an immune response to the protein, has Her-2/neu extracellular domain fused to Her-2/neu intracellular or phosphorylation domain.

Claim 68: Fig 7: 141pp: English.

The present sequence is that of human Her-2/neu (p185 glycoprotein or c-erbB2), an oncogenic self-protein and target for anti-cancer vaccines. The Her-2/neu gene is amplified and p185 is overexpressed in a variety of cancers, including breast, ovarian, colon, lung and prostate cancer. Her-2/neu is a member of the tyrosine kinase family of receptor-like glycoproteins. It comprises an extracellular domain with homology to the epidermal growth factor receptor (EGFR), a highly hydrophobic transmembrane domain and a C-terminal intracellular domain that also shows homology to EGFR. Its overexpression correlates with a poor prognosis in breast and ovarian cancers. The invention provides Her-2/neu fusion proteins, nucleic acids encoding them, viral vectors, and vaccines comprising the fusion proteins or nucleic acid molecules. In preferred fusion proteins, the extracellular domain of a Her-2/neu protein is fused to a Her-2/neu intracellular domain or phosphorylation domain (or its DeltapD fragment). An immune response to Her-2/neu protein is elicited or enhanced by administering the fusion protein in the form of a vaccine, or by transfecting cells of an animal *ex vivo* with a nucleic acid encoding the fusion protein, and delivering the transfected cells to the animal. The fusion proteins, nucleic acids, and isolated specific T-cells are useful for inhibiting the development of a cancer, especially breast, ovarian, colon, lung or prostate cancer in a patient. T cells that specifically react with a Her-2/neu fusion protein can be used to remove tumour cells from a sample in order to inhibit the development of cancer in a patient.

Sequence 1255 AA:

ery Match	98.0%;	Score 6709;	DB 23;	Length 1255;
st Local Similarity	98.0%;	Pred. No. 0;		

1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPEETHLDMLRLHYQGCGVVOGNL 60
|||||
1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPEETHLDMLRLHYQGCGVVOGNL 60

QY 61 ELTYLPTNASLSFLQDIOEQVGVLYIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEQVGVLYIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDF---NNFTV 177
DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKN--- 177
QY 178 SFMLRPVKVSASHLE-NRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 236
DB 178 -----QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGP 230
QY 237 LPTDCHEQCAACTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMNPBGRT 296
DB 231 LPTDCHEQCAACTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFESMNPBGRT 290
QY 297 FGASCVTACPYNLTSDVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGMEHL 356
DB 291 FGASCVTACPYNLTSDVSGCTLVCPHNOEVAEDGTQRCCKSPCARVCYGLGMEHL 350
QY 357 REVRVTSANIQSFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 416
DB 351 REVRVTSANIQSFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSITLQGLGISWGLSLRSLRELGLALIIH 476
DB 411 YISAWPDSLPLDSVFQNLQVIRGRIILHNGAYSITLQGLGISWGLSLRSLRELGLALIIH 470
QY 477 NTHLCFVHTVPMDLFRNPHOALLHTANRPEDSCVGEGLACHOLCARGHCWGPGTQCVN 536
DB 471 NTHLCFVHTVPMDLFRNPHOALLHTANRPEDSCVGEGLACHOLCARGHCWGPGTQCVN 530
QY 537 CSQFLRGQECVEECRVLQGLPREYVVARHCLPCHPECQPNQSVTCFGEADOCVACAHY 596
DB 531 CSQFLRGQECVEECRVLQGLPREYVVARHCLPCHPECQPNQSVTCFGEADOCVACAHY 590
QY 597 KDPFPCVACPSGVKPDLSMPYIWKFPDEBACQPCINCTHSCVDLDDKGCFAERASP 656
DB 591 KDPFPCVACPSGVKPDLSMPYIWKFPDEBACQPCINCTHSCVDLDDKGCFAERASP 650
QY 657 LTSIVSAVGVILLVVLGVVFGILIKRROQKIRKYMTRRLLOETELVEPLTPSGAMPNQA 716
DB 651 LTSIVSAVGVILLVVLGVVFGILIKRROQKIRKYMTRRLLOETELVEPLTPSGAMPNQA 710
QY 717 QMRILKETELRKVKVLSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDE 776
DB 711 QMRILKETELRKVKVLSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKELDE 770
QY 777 AYVMAGVGSPPYVSRLLGICLTSTVQLVQLMPYGCLLDHVRENRLGSDLLNWCWQIA 836
DB 771 AYVMAGVGSPPYVSRLLGICLTSTVQLVQLMPYGCLLDHVRENRLGSDLLNWCWQIA 830
QY 837 KGMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 896
DB 831 KGMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGGKVPKWWA 890
QY 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 956
DB 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 950
QY 957 VYMIWVKWMIIDSECPRELVSEFSRMARDPQRFVITQNEIDLGPASPLDSTFYRSLLE 1016
DB 951 VYMIWVKWMIIDSECPRELVSEFSRMARDPQRFVITQNEIDLGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSTSTSGGDLTLGLEPSEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSTSTSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130
QY 1137 LTCSPQEVYVQPDVPRQPPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVFAGGA 1196

DB 1131 LTCSPQEVYVQPDVPRQPPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVFAGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYDQDPPERCAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYDQDPPERCAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255
RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML; acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL; chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS; Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide encoding the polypeptide, or antigen presenting cells expressing the polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of haematological malignancy in a patient by administering a polypeptide comprising an immunogenic portion of Her-2/neu or a polynucleotide encoding the polypeptide. Antigen presenting cells that express the protein can also be administered. The sequences are used for inhibiting development of haematological malignancy such as acute myelogenous leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6709; DB 23; Length 1255;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;
QY 1 MELALCRWGLLALLPPGAASVQCTGDMKRLPASPETHLMLRHLVQGCQVQGNL 60
DB 1 MELALCRWGLLALLPPGAASVQCTGDMKRLPASPETHLMLRHLVQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEQVGVLYIAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

|||||
61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||||
121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDF---NNFTV 177
|||||
121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDIFHNKN--- 177
|||||
178 SFWLVRPKVSASHLE-NRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKGP 236
::: : : :
178 -----QLALTLDITNRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKGP 230
|||||
237 LPTDCCHEQCAAGCTGPKHSOCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
|||||
231 LPTDCCHEQCAAGCTGPKHSOCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
|||||
297 FGASCVTACPNYLSLTDVSGCTLVCPLNHQVTAEDGTQRCCKSKPCARVCYGLGMEHL 356
|||||
291 FGASCVTACPNYLSLTDVSGCTLVCPLNHQVTAEDGTQRCCKSKPCARVCYGLGMEHL 350
|||||
357 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 416
|||||
351 REVRAVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
|||||
417 YISAWPDSLPLDSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIH 476
|||||
411 YISAWPDSLPLDSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIH 470
|||||
477 NTHLCFVHTVPMWDLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCGPPTQCVN 536
|||||
471 NTHLCFVHTVPMWDLFRNPHQALLHTANRDECEVGEGLACHOLCARGHCGPPTQCVN 530
|||||
537 CSQFLRGQECVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCACAHY 596
|||||
531 CSQFLRGQECVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFGEADQCACAHY 590
|||||
597 KDPFFCVARCSGVKPDLSYMPIWKFDPBEGACQPCINCHSCVDLDDKCPABEQRAS 656
|||||
591 KDPFFCVARCSGVKPDLSYMPIWKFDPBEGACQPCINCHSCVDLDDKCPABEQRAS 650
|||||
657 LTSIVSAVVGILLVVLGWFGILIKRQKIRKVTMRLLQETELVEPLTPSGAMPNOA 716
|||||
651 LTSIVSAVVGILLVVLGWFGILIKRQKIRKVTMRLLQETELVEPLTPSGAMPNOA 710
|||||
717 QMRILKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776
|||||
711 QMRILKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 770
|||||
777 AYVMAGVSPYVSRLLGLCLTSTVOLVTQLMPYGCCLLDHVRENRRGLSGQDLNWCQIA 836
|||||
771 AYVMAGVSPYVSRLLGLCLTSTVOLVTQLMPYGCCLLDHVRENRRGLSGQDLNWCQIA 830
|||||
837 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWA 896
|||||
831 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWA 890
|||||
897 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
|||||
891 LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
|||||
957 YVIMVVKWMTDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFVRSLE 1016
|||||
951 YVIMVVKWMTDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFVRSLE 1010
|||||
1017 DDDMGDLVDAEYELVPQGFCCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEBEA 1076
|||||
1011 DDDMGDLVDAEYELVPQGFCCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEBEA 1070
|||||
1077 PRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAP 1136
|||||
1071 PRSLPASEGAGSDVFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAP 1130
|||||
1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGGA 1196
|||||

Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAGGA 1190
Qy 1197 VENPEYLTPOGGAAPQPHPPAFSPAFDNLYWQDDPPERGAPPSTFKGTPTAENPEYLG 1256
|||||
Db 1191 VENPEYLTPOGGAAPQPHPPAFSPAFDNLYWQDDPPERGAPPSTFKGTPTAENPEYLG 1250
|||||
Qy 1257 LDVPV 1261
|||||
Db 1251 LDVPV 1255
|||||
RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
FN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston Lb, Huston JS, Oppermann H, Ring DB;
DR WPI; 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
Query Match 97.3%; Score 6666; DB 14; Length 1433;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1233; Conservative 9; Mismatches 9; Indels 14; Gaps 3;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHDMLRHLGYCQVVGQNL 60
|||||
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHDMLRHLGYCQVVGQNL 60
|||||
Qy 61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||||
Db 61 ELTYLPTNASLFLQDIQEVGYVLI AHNVQVRVPLQRLRIVRGTQLFEDNYALAVLDNG 120
|||||
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDF---NNFTV 177
|||||
Db 121 DPLNNTTPTVTCASPGGLRELQRLSITELTKGGVLIQRNPOLCYQDTILWKDIFHNKN--- 177
|||||
Qy 178 SFWLVRPKVSASHLE-NRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKGP 236
::: : : :
Db 178 -----QLALTLDITNRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKGP 230
|||||

Search completed: July 22, 2003, 08:40:44
Job time : 43.1589 secs

```
OY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPGRYT 296
DB |||||
DB 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPGRYT 290
OY 297 FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQCEKCKPCARVCYGLGWEHL 356
DB |||||
DB 291 FGASCVTACPYNYLSTDVGSCTLVCPHNOEVTAEQCEKCKPCARVCYGLGWEHL 350
OY 357 REVRAVTSANIQEFACKKIFGSLAPFESFDGDPASNTAPLQPEQLQVFETLEEITGYL 416
DB |||||
DB 351 REVRAVTSANIQEFACKKIFGSLAPFESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
OY 417 YISAWPDSLPDLVSFONLQVIRGRILHNGAYSITLQGLGSIWGLRSLRELGSGLAIHH 476
DB |||||
DB 411 YISAWPDSLPDLVSFONLQVIRGRILHNGAYSITLQGLGSIWGLRSLRELGSGLAIHH 470
OY 477 NTHLCFVHTVPWDOLFNRPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVN 536
DB |||||
DB 471 NTHLSFVHTVPWDOLFNRPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVN 530
OY 537 CSQFLRGQCBECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHY 596
DB |||||
DB 531 CSQFLRGQCBECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVACAHY 590
OY 597 KOPPPFCVAPCPGSKVDLSYMPIWKFPDEGACQPCPNCTHSCVDLDDKGCAPABORASP 656
DB |||||
DB 591 KOPPPFCVAPCPGSKVDLSYMPIWKFPDEGACQPCPNCTHSCVDLDDKGCAPABORASP 650
OY 657 LTSISAVVGILLVVVLGVFGILIKRROQKIRKYMRRLLQETELVEPLTPSGAMPNQA 716
DB |||||
DB 651 LTSISAVVGILLVVVLGVFGILIKRROQKIRKYMRRLLQETELVEPLTPSGAMPNQA 710
OY 717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPLANKEILDE 776
DB |||||
DB 711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPLANKEILDE 770
OY 777 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRRGLGSODLLNWCQIA 836
DB |||||
DB 771 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRENRRGLGSODLLNWCQIA 830
OY 837 KGMVYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896
DB |||||
DB 831 KGMVYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890
OY 897 LESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICTID 956
DB |||||
DB 891 LESILRRRTHQSDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPPICTID 950
OY 957 VYMIWVKWMDSECRPRELVSERSMARDPQRFVTONEDLGPASPLDSTFYRSLLE 1016
DB |||||
DB 951 VYMIWVKWMDSECRPRELVSERSMARDPQRFVTONEDLGPASPLDSTFYRSLLE 1010
OY 1017 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGMVHHRSSSTRSGGDLTLGLEPSEEA 1076
DB |||||
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGMVHHRSSSTRSGGDLTLGLEPSEEA 1070
OY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAP 1136
DB |||||
DB 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAP 1130
OY 1137 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARFAGATLERAKTSLSPKNGVVKDVFAGGA 1196
DB |||||
DB 1131 LTCSPQPEYVNOQDVRPQPPSPREGPLPAARFAGATLERAKTSLSPKNGVVKDVFAGGA 1190
OY 1197 VENPEYLTPOGGAAPQPHPPPAFPAFDNLYYWDODPPERGAPPSTFKGTPTAENPEYLG 1256
DB |||||
DB 1191 VENPEYLTPOGGAAPQPHPPPAFPAFDNLYYWDODPPERGAPPSTFKGTPTAENPEYLG 1250
OY 1257 LDVPV 1261
DB |||||
DB 1251 LDVPV 1255
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-210-224-14
Perfect score: 6827
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGDPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: Pirl:.*
2: Pirl2:.*
3: Pirl3:.*
4: Pirl4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6684	97.9	1255	1 A24571	protein-tyrosine k
2	5870.5	86.0	1254	2 I48161	p-185 precursor
3	5866	85.9	1260	1 TVRTNU	protein-tyrosine k
4	3054	44.7	1210	1 GOHUE	epidermal growth f
5	3030	44.4	1210	2 A53183	epidermal growth f
6	3009.5	44.1	1223	1 TVCHLV	epidermal growth f
7	2905.5	42.6	1308	2 A47253	epidermal growth f
8	2602	38.1	1166	1 S06142	protein-tyrosine k
9	2333.5	34.2	1342	2 A36223	kinase-related tra
10	2248.5	32.9	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUHV	protein-tyrosine k
13	1647	24.1	544	2 S35745	protein-tyrosine k
14	1640	24.0	545	2 S00727	protein-tyrosine k
15	1623	23.8	540	2 B44776	kinase-related tra
16	1621	23.7	540	1 TVFVEB	protein-tyrosine k
17	1561.5	22.9	1330	1 GQFPE	epidermal growth f
18	1432	20.8	644	2 A36325	epidermal growth f
19	1251	18.3	1323	2 E88257	protein let-23 [im
20	1251	18.3	1374	2 S70712	protein-tyrosine k
21	1171	17.2	1369	2 S70713	protein-tyrosine k
22	1142	16.7	1717	1 A45558	epidermal growth f
23	1041	15.2	527	2 A42032	epidermal growth f
24	906.5	13.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	680	10.0	1382	1 INHUR	insulin receptor p
28	671	9.8	1363	2 T43220	insulin-like growt
29	671	9.8	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A44188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, I.
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:3003577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epic
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:G183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcription
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

ALIGNMENTS

A:Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NGL; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A:Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
inase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-125/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carbonylate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

	Query Match	97.9%	Score 6684;	DB 1;	Length 1255;
	Best Local Similarity	97.6%	Pred. No. 3.1e-279;		
	Matches 1239;	Conservative	2;	Mismatches 6;	Indels 22; Gaps 2;
Qy	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMLKRLPAS	PETHLDMLRHLYQGCVVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMLKRLPAS	PETHLDMLRHLYQGCVVQGNL	60
Qy	61	ELTYLPTNASLSFLODIOEQGYVLI	AHNQVRQVPLORLIRVGTQLFEDNYALAVLDNG	120	
Db	61	ELTYLPTNASLSFLODIOEQGYVLI	AHNQVRQVPLORLIRVGTQLFEDNYALAVLDNG	120	
Qy	121	DPLNNTTPTVGASPGGLRELQRLSL	TEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA	180	
Db	121	DPLNNTTPTVGASPGGLRELQRLSL	TEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA	180	
Qy	181	LTLIDTNRSRACHPCSPMKGSR	CWGESSEDCQSLTRTVCAAGCARCKG	BPLPNNFTVSF	240
Db	181	LTLIDTNRSRACHPCSPMKGSR	CWGESSEDCQSLTRTVCAAGCARCKG	BPLPNNFTVSF	240
Qy	241	WLRVPKVSASHLE	-----KHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPE	232	
Db	233	-----TDCHEQCAAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPE	286		
Qy	293	GRYTFGASCVTACPYNYLSTDVGSCT	ILVCLPHNQVETAEDGTORCEKCKSPCARVCYGLG	352	
Db	287	GRYTFGASCVTACPYNYLSTDVGSCT	ILVCLPHNQVETAEDGTORCEKCKSPCARVCYGLG	346	
Qy	353	MEHLREVRVTSANIQSFAGCKITFGSLA	FLPESFDGDPASNTAPLOPEQLQVFTELEEI	412	
Db	347	MEHLREVRVTSANIQSFAGCKITFGSLA	FLPESFDGDPASNTAPLOPEQLQVFTELEEI	406	
Qy	413	TGYLIYSAMPDLSLDLSVFQNLQVIR	GRILHNGAYSILTQGLGISTWGLSLRELGSGLA	472	
Db	407	TGYLIYSAMPDLSLDLSVFQNLQVIR	GRILHNGAYSILTQGLGISTWGLSLRELGSGLA	466	
Qy	473	LIHNHNLFCFVHTVPWDLFRNPHQALL	TANRPDEBCVGEGLACHOLCARGHCWGPPT	532	
Db	467	LIHNHNLFCFVHTVPWDLFRNPHQALL	TANRPDEBCVGEGLACHOLCARGHCWGPPT	526	
Qy	533	QCVNCSOFLRGQECVEECRVLQGL	PREYVVARHCLPCHPECPQNGSVTCFGEAOCVA	592	
Db	527	QCVNCSOFLRGQECVEECRVLQGL	PREYVVARHCLPCHPECPQNGSVTCFGEAOCVA	586	
Qy	593	CAHYKDPFFCVARCPGKVPDL	SYMPITWKFPEDEGACQPCPINCTHSCVDLDKGCFAEQ	652	

```
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 1 MELAAWCGWGLLLALLSPGASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTANATLSFLQDIOEVGYVLIAHNQVQVPLQRLRVRGTLQFEDNYALAVLNR 120
Qy 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFHNKQOLA 180
Db 121 DPLDNVTATGRTPEGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFHNKQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCSWGSESDCQSLTRTV-----C 220
Db 181 PVDIDTNRSRACPPCAPACKDNHCWGSPEDCQTLTGTTIAPRAVPAARLARLPTDCCHQOC 240
Qy 221 AGGCARCKGPLPFNNFTVSFWLRVPKVSASHLEKHSCLACLHNHSGICELHCPALVTY 280
Db 241 AAGCT---GP-----KHSDCLACLHNHSGICELHCPALVTY 274
Qy 281 NTDTFESMPNPEGRYTFGASCVTACPNYLVSTDVSGSCTLVCPLNHQNVEVTAEDGTQRCCK 340
Db 275 NTDTFESMPNPEGRYTFGASCVTTCPNYLVSTEVSGSCTLVCPLNHQNVEVTAEDGTQRCCK 334
Qy 341 SKPCARVCYGLGMEHLREVRVAVTSANTQIEFAGCKKIFGSLAFLESFDDGPASNTAPLQP 400
Db 335 SKSCARVCYGLGMEHLRGARAITSANIQEFAGCKKIFGSLAFLESFDDGNSSGTAPLTP 394
Qy 401 EQLQVFETLEBITGYLYISAMPDSILPDLVSFQNLQVIRGRILHNGAYSLTLOGLISHLG 460
Db 395 EQLQVFETLEBITGYLYISAMPDSLHDSVLFQNLQVIRGRILHNGAYSLAQGLGIRWLG 454
Qy 461 LRSRELGSGLALIHNTLHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGBGLACHOL 520
Db 455 LRSRELGSGLVLIHRNTHLCFVHTVPMQDLFRNPHQALLHSGNPESECGLKDFACVPL 514
Qy 521 CARGHCWGPGTQCNSQFIRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSV 580
Db 515 CAHGHCWGPGTQCNSHFIRGQECVKECRVWKGLPREYVNGKHLCPCHPECPQNSTE 574
Qy 581 TCFPEADOCVACAHYKDPFCVARGCPGVKPDLSYMPIKWFPEEGACQCPNCTHSC 640
Db 575 TCTGSEADQCTACPHYKDSPPFCVARGCPGVKPDLSYMPIKWFPEEGMCQPCPNCNTHSC 634
Qy 641 VLDLDDKCPAQRASPLTSIVSAVVGLLVVGLVGVFGLIKRRQKIRKYMRLLOET 700
Db 635 VLDLDDKCPAQRASPLTSIIATVVGILLFLVIGVVGILLIKRRQKIRKYMRLLOET 694
Qy 701 ELVEPLTPSGAMPNQAQNRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKV 760
Db 695 ELVEPLTPSGAMPNQAQNRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKV 754
Qy 761 LRENTSPKANKEILDEAVVWAGVSPYVSRLLGICLTSTVOLVTLQMPYGCGLLDHVRNR 820
Db 755 LRENTSPKANKEILDEAVVWAGVSPYVSRLLGICLTSTVOLVTLQMPYGCGLLDHVRNR 814
Qy 821 GRLSQDILLNCMQIAKGMYSLEVDVLRHDLAARNVLKSPNPKVITDFGLARLLDIDE 880
Db 815 GRLSQDILLNCWQIAKGMYSLEVDVLRHDLAARNVLKSPNPKVITDFGLARLLDIDE 874
Qy 881 TEYHADGCKVPIKWMALLESILRRRFTHQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDL 940
Db 875 TEYHADGCKVPIKWMALLESILRRRFTHQSDVWSYGVTVWELMTGAKPYDGIPIAREIPDL 934
Qy 941 LEKGERLPQPPICITIDVYVIMVKCWMIDSECRPRFELVSEFSRWAPDQRFVUIONEDL 1000
Db 935 LEKGERLPQPPICITIDVYVIMVKCWMIDSECRPRFELVSEFSRWAPDQRFVUIONEDL 994
Qy 1001 GPASPLDSTFYRSLLEDDMGDLVDABEYLVPOQGFCCPDPAAGAGGVVHRRHSSSTRS 1060
Db 995 GPSSPLDSTFYRSLLEDDMGDLVDABEYLVPOQGFCCPDPAAGAGSTAHRHSSSTRS 1054
Qy 1061 GGGDLTLGLEPSEEEAPRSLAPSEGAGSVDFDGLGMAAGKGLQSLTPHDPSPLQRYSE 1120
```

```
Db 1055 GGGELTLGMEPSGEEPPRSLAPSEGAGSVDFEGLMGATKGPQISPRDLSPLQRYSE 1114
Qy 1121 DPTVPLSETGYVAPLTCSPQPEVYNQDPVVRPQPPSPREGPLPAARPAGATLERAKTUS 1180
Db 1115 DPTLPLPETDGYVAPLACSPQPEVYNQDPVVRPQPPSPREGPLPVRPAGATLERPKTUS 1174
Qy 1181 PGKGVKVDVFAFGGAVENPEYLTPQGGAAAPQHPHPPAFSPAFDNLXYWQDPPPPGAPP 1240
Db 1175 PGKGVKVDVTFGGAVENPEYLVPRGGSASQPH-PPALCPAFDNLXYWQDPPSERGSP 1233
Qy 1241 STFKGTPTAENPEYLGLDVVP 1261
Db 1234 NTFEGTPTAENPEYLGLDVVP 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Batgmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:956746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carlinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a
2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 637-663, 'V', 665-702 <MAS>
A:Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
A:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; signal sequence; glycoprotein; phosphoprotein; phospho-
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase new #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 85.9%; Score 5866; DB 1; Length 1260;
Best Local Similarity 85.7%; Pred. No. 3,Se-244;
Matches 1089; Conservative 51; Mismatches 107; Indels 24; Gaps 4;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 4 MELAAWCGWGLLLALLSPGASCTGCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 63
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVRGTLQFEDNYALAVLNG 120
Db 64 ELTYVPAANASLSFLQDIOEVGYVLIAHNQVQVPLQRLRVRGTLQFEDNYALAVLNR 123
Qy 121 DPLNNTTPTVTCASPGGLREQLRLSLTEILKGGVLIQRNPQLCYODTILWKDIFHNKQOLA 179
Db 124 DPQDNVAASFTGRTPEGLREQLRLSLTEILKGGVLIQRNPQLCYODVWLKDVFRKNNQL 183
Qy 180 ALTLIDTNRSRACHPCSPMKSGRCSWGSESDCQSLTRTVTCAGGCARCKGPLPNNFTVS 239
Db 184 APVIDTNRSRACPPCAPACKDNHCWGSPEDCQTLTGTTCTSCARCKGRLP----- 236
```



```

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lab, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mod
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #E
F:192,650/Binding site: carboxylate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.1%; Score 3009.5; DB 1; Length 1223;
Best Local Similarity 47.3%; Pred. No. 86-122;
Matches 620; Conservative 176; Mismatches 348; Indels 167; Gaps 27;
QY 8 RWGLLLALLPPCAA-----STQVCTGTDMLRLPASPTHLDMLRLHYQGCVVQGNLE 61
DB 13 RGAALVLLVLLGVALCSAVEEKVCQGTNNKLTQLGHVEDHFTSLQRMVNNCEVLSNLE 72
QY 62 LTYLPTNASLFLQDIQVQGYLVLIHNOVROVPLQRLRIVRGTLFEDNYALAVLDNGD 121
DB 73 ITVVEHNRDLTLKTIQEVAGVLIALNMVDVPLENQLIIRGNVLYDSFALVLSNTH 132
QY 122 PLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFPHKNQAL 181
DB 133 -MNKTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPKCKGRGSESDCSLTTRTVACGGCA-RCKGPLPFNNFTVS 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVCAQQCSGRCKGVPSD----- 237
QY 240 FWLRLVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVYNTDTFESMNP 291
DB 238 -----CCHNCAAGCTGPRESDCLACRFRDATTCKTCCPLVLYNPTYQMDVNP 288
QY 292 EGRYTFGASCVTACPYNYLSTDVSGCTLVCPILHNOEVTAEADGTQRCCKSPCARVCYGL 351
DB 289 EGKYSFGATVRECPHNVVTDHGSVRSCTNDDTYEV-EENGVRCKCKCDGLCSKVCNGI 347
QY 352 GMEHLREVRVATSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVLPETLEE 411

```

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culicou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/pi80erbB4, a fourth member of the epiderm

A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <P10>
A;Cross-references: GB:J07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;774-732/Region: protein kinase ATP-binding motif

Query Match 42.6%; Score 2905.5; DB 2; Length 1308;
Best Local Similarity 44.7%; Pred. No. 2.4e-117;
Matches 608; Conservative 181; Mismatches 376; Indels 195; Gaps 31;
Qy 9 WGLLLALLPGAA-----STQCTGTDMLKRLPASPETHDMLRHLGYGCVVQGNLBTY 64
Db WVVSVLLVAAGTVQPSDSQSVACAGTENKLSLSLEQQYRALRYKYEENVMGNLEITS 67
Qy 65 LPTNASLSFLQDIOEVGYVLIHNOVQVPLQRLRIVRGTLQFEDNVALAVLDGDLN 124
Db 68 IEHRDLISLRSREVTVGYVLVALNQRYLPLENLRIRGTLKYEDRYALAIFLNYRKDG 127
Qy 125 NTPVTGASPGGRELQRLSLTEILKGGVLIQRPOLCYQDTILKWDIFHKNNQALTLI 184
Db 128 NF-----GLQELGLKNLTELNGVGVYDQNKFLCYADTIHWQDIVRNPWPSNLTV 178
Qy 185 DTNRSRACHPCSPMKGSRGWGSESDCOSLTRVACAGC-ARCKGPLPNNFTVSWLR 243
Db 179 STNGSSCGRCHKSCG-RCWGPTENHCOTLTRVCAEQDGRCYGY-----VS----- 227
Qy 244 VPKVSASHLE-----KHSDDLACLHFNHSGICELHCPALVTYNTDTFESMNPBGY 295
Db 228 ----DCCHREACGCGSPKOTDCFACMNFDSGACVTCQPTFYNYNTTQLEHFNKAY 283
Qy 296 TFGASCVTACPYNYLTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCVYGLQMEH 355
Db 284 TYGAFVCVKCPHNFV-VDSSSVCRAKPSRMEV-EENGIRKCKPCTDICPKACDGIOTGS 341
Qy 356 LREVRVTSANIOEFACCKIFGLSLAPESFGDPSANTAPLOPEOLQVETLEBITGY 415
Db 342 LMSAQTVDSNIDKFNCTKINGNLIFLVTGIHGDYPNAIEADPEKLNFRVREITGF 401
Qy 416 LYSAMPDLSLDSVFQNLQVIRGIRLHNGAYSILTQGLGTSWLGSLRSLRSLGSLALIH 475
Db 402 LNIQSMPPNMTDSVFSNLVTIGRVLYSGLSLILKQGITSLQFOSLKEISGNVIIT 461
Qy 476 HNTLCFVHTVPDQLPRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCWGPGTQCV 535
Db 462 DNSNLCVYHTINWTLTSTINQIRIVIRDNRAENCTAEGVMVCHLSSDGCWGPGDQCL 521
Qy 536 NCSQFLRGQCEVECRVLQGLPREYVNAHCLPCHPECOP-ONGSVTCFGEADQCACA 594
Db 522 SCRRFSGRICISCSNLDYGEFEFENGSI-CVECDPQCEKMEGDLTLTCHGPGDNCCKS 581
Qy 595 HYKDPPEFCVACRSPGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGC----- 648
Db 582 HFKDGPNCKECPDGLQANSF--IFKYADPDRECHPCNCTQGCGNPTSHDCIYYPWT 639
Qy 649 -----PAEQRASPLTISVAVV-GILLVVLGVVFGVILIKRQOKIRKYTRRLLOTEL 702
Db 640 GHSTLPQAHAR-TPL--IAAGVIGGLFVLIVGLTFVAVYVRKSGIK-KKRALRRFL-ETEL 694
Qy 703 VEPLTPSGAMPNOAMRILKETELRKVKVLSGAGFVYKGIWIPDGENVKIPVAIKVLR 762
Db 695 VEPLTPSGTAPNAQARILKETELRKVKVLSGAGFVYKGIWIPDEGETKIPVAIKILN 754
Qy 763 ENTSPKANKEILDEAYVMAGVSPYSRLIGLICLTSTVQLVQTLMPYGCILLDHVRENRR 822
Db 755 ETTGPKRANVEFMDEALIMASMDPHLVRLLGVCLSPITQLVQLMPHGCLLEYVHEKDN 814
Qy 823 LGSQDLLNWCQIAKGMVSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDDETE 882

Db 815 IGSQDLLNWCQIAKGMVSYLEDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDDETE 874
Qy 883 YHADGGKVPKWMMALESILRRRTHQSDVWSYGVVWELMTFGAKPYDCIPAREIPDILLE 942
Db 875 YNADGGKMPKWMMALECIHYRKETHQSDVWSYGVVWELMTFGGKPYDGIPTREIPDILLE 934
Qy 943 KGERLPQPPCTTIDVVMVVKCMIDSECRPRRRELVSERPMARDPQRFVVQIENED-LG 1001
Db 935 KGERLPQPPCTTIDVVMVVKCMIDADSRPKFKELAAEFRRMARDPQRYLVLIQODDRMK 994
Qy 1002 PASPLDSTFYRSLLEDDDDMGDLVDAAEYLVPPQGFCCPDAPAGAGMVHRRHSSTRSG 1061
Db 995 LPSNDISKFFNLLDEEDLEMDMAEYLVLP-QAFNIPPP-----IVTSRARDISNRS- 1046
Qy 1062 GGDLTIGLSESEAPRS-----PLAP-SEGAGSDVP 1092
Db 1047 ----EIGHSPPPAYTPMSGNQFYVYRQDGAEGVSVYPYRAPTSTIPEAPVAQATAEIF 1102
Qy 1093 DGDLGMAAGKGLQSLPTHDPSPLOQYSEDTPTVPLPS-----ETDGYVAPLTCSPQPEY 1145
Db 1103 DSCCNGTLRKVPAPHVQESSTQRYSDADPTVPAPERSPRGELDEEGYWTMRDKPKQEY 1162
Qy 1146 VNQPDVVRQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVGVKDVFAFGGAVENPEYLTP 1205
Db 1163 LNPVE-----ENPFVSR-----KNGDLQ-----ALDNPEYHNA 1191
Qy 1206 QCGAAPQPHPPA-----FSPAFDNLVYWDQ 1231
Db 1192 SNG-----PKAEDVEYNEPLYLNTFANTLGKAEYLNILSMPEKAKKAFDNPDYWNH 1245
Qy 1232 DPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1246 SLPRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8

S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transform
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Mittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe
Nature 341, 415-421, 1999
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R;Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Introns: 872/3; 898/1; 947/3; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif
Query Match 38.1%; Score 2602; DB 1; Length 1166;
Best Local Similarity 44.5%; Pred. No. 2.2e-104;

Marches	567;	Conservative	164;	Mismatches	396;	Indels	148;	Gaps	30;
Qy	4	AALCRWGLLLALLPPGA	AST----	QVCTGTDMKLR	PASPETHLDMRLH	LVGGQV	VOGN	59	
Db	8	AALLQ--LLL	LSIRCCSTDP	DRKVC	GTSSNQ	MTM---	LDNHV	LKMKMYSGCN	62
Qy	60	LELTYPLNASLS	FLQDIOEQVGY	VLIAHNOVRQ	PLORLRI	VRGTQ	L	PENYALAVLDN	119
Db	63	LEITYTOENQDLS	FLQSIQEVGY	VLIAHNEVSTI	PLNLRIR	IGONLYE	GNFTLL	VMSN	122
Qy	120	GDPLNNTT	PVTGASPG	RLRLQLAS	LTEILKGG	VLIORN	POLCYQ	DTILWKD	179
Db	123	YQK-NPSSP--	-DVYQVGLK	QLQLNSL	TEILSGG	VKSHNP	LLCNVETI	NWWDI	179
Qy	180	ALTIDTNRSR	ACHPCSPMK	SGSRCSW	GESSDC	OSLTR	TVTCAGC	-ARCKG	236
Db	180	TWNLI	PHAFERQ	CKDCDHG	CYVNGSW	APGHC	OKFTKLL	CAEQNRR	239
Qy	237	TVSFWR	VRPKVSASH	LKHS	DLACLHP	NSGICEL	HC	PALVT	290
Db	240	HCAGCTG	PRA-----	TDCLAC	RDFND	DGTC	KDTPPK	YDI	291
Qy	297	FGASC	VTACP	NYLST	DVGS	CTLV	CPHLNHQ	EVTAEDG	356
Db	292	FGAAC	KECP	SNYYVTE	-GACV	RS	SAGL	EV	349
Qy	357	REVR	AVTS	ANIQE	FAGCK	KIFG	S	LAFLPES	416
Db	350	SNTI	AVNST	NI	RSFS	NCTK	INGDI	IILNRNS	409
Qy	417	YIS	APD	SLP	DS	LVFQ	NI	QVTRGR	475
Db	410	VIM	W	PEN	MT	SL	VS	QNL	469
Qy	476	HNTH	LCF	VHT	PM	D	LPEN	PHO	535
Db	470	NLQ	LR	YANT	IN	WR	AL	FR	521
Qy	536	NCS	O	FLR	GOE	CB	EC	RV	595
Db	522	SCL	H	VD	R	G	R	C	581
Qy	596	YK	D	P	P	F	C	V	655
Db	582	FQD	Q	P	Q	I	P	R	639
Qy	656	PLT	S	I	S	A	V	G	715
Db	640	HSS	I	A	V	G	L	V	698
Qy	716	AQ	M	I	K	E	T	E	775
Db	699	AFL	I	L	K	E	T	E	758
Qy	776	EAY	M	A	G	V	G	S	835
Db	759	EAY	M	A	V	A	S	V	818
Qy	836	AKG	S	L	E	D	V	L	895
Db	819	AKG	M	N	Y	L	E	R	878
Qy	896	ALES	I	L	R	R	R	F	955
Db	879	ALES	I	L	Q	W	T	H	938
Qy	956	DVY	M	I	M	V	K	M	1015
Db	939	EYV	M	I	L	K	M	D	995
Qy	1016	EDD	M	G	L	V	D	A	1075
Db	996	SSD	-	-	-	-	-	-	1019

Qy 1076 APRSLAPFSEGAGSDVFDGDLGMGAQLQSPLTHDPSPLQRYSXEDPTV-PLPSETDGTV 1134
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1020 --EPCIPETGH-----PVRENSITLRNISDPTQNALDKLDGH- 1055
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1135 APLTCSPOPEYVNPQDRVRQP-----PSPRE-----GLP-AARPAGATLERAKTLSP 1181
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1056 -----EYNYPQGETSSRLSDTYNPNYEDLTDCMGVPVSJSSQAETNFSRPEYLNT 1106
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1182 GKNGVKVDVFARFGGAVENPEYLTPOGGAAPQHPPPAFPAFNLYYWDQDPPIRGAPP 1241
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1107 NQNSI---PLVSSSGMDDPDY---QAG-----YQAAF-----LPOTGALTG 1141
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 1242 TFKGTPTAENPEYLG 1256
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1142 NGMFAPAENLEYLG 1156
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, P.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor receptor
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:I19880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
C;Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 34.2%; Score 2333.5; DB 2; Length 1342;
Best Local Similarity 39.7%; Pred. No. 8.1e-93;
Matches 526; Conservative 188; Mismatches 460; Indels 151; Gaps 34;

Qy 10 GLLLALLPPGAA--STQVCTGTDMLKLRLPASPTHLDMLRHLYQGCGVOGNLELTYLPT 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 11 GLLFSLARGSEVGNSQAVCPGTLNGLSVTGDAENQYTLYKYERCEVVMGNLEIVLTGH 70
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 68 NASISFLQDIQEVOGYVLIAHNOVQVPLRIVRGTOLFFEDNYALAVLDNGDPLNNTT 127
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 71 NADLSFLOWIREVTGVLVVANNEFTLPLPNLRVVRGTQYIDKGKFAIFYN-----LNynt 125
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 128 PVTGASPGGLRELQRLSUTEILKGGVLIQRNPQCXYQDTILMKDIFHKNNQLAULTIDTN 187
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 126 ----NSSHALRQLRTLQTLEILSGGVYIEKNKLCMDITDMREDIVRDND---AEIVVKD 178
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 188 RSRACHPCSPCKGRCHGESEDQSLTRTCVAGGC-ARCXGPPLFPNNFTVFWLRVPK 246
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 179 NGRSCPPCHEVCYG-RCWGPSGEDCQTLTKTICAPQCNHGCFGNP----- 223
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 247 VSASHLE-----KHSDCLACLHNHSGIGELCHPALVTYNTDTFESMPNPEGRYTFG 298
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```
Db 224 NQCCHDEACGCGSPQDTCFACRHFNDGACVPRCPQLVYNKLTFLQEPNPHTKYQG 283
Qy 299 ASCVTACPNYLSLTDVSGSCTLVCPHUNQEVTAEDGTQRCCKSPCARVCVGLGMEHLRE 358
Db 284 GVCVASCPHNFV-VDTQSCVRACPPDRKMEVD-KNGLKNCEPCGGLCPKACCEGTSG--SR 339
Qy 359 VRVTSANIOBFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLEETGYLYI 418
Db 340 FQTVDSNIDCFVNCTKILGNLDFLTGLNGDPWHKIPALDPEKLNVPRTVREITGYLNI 399
Qy 419 SAWPDSLPLSVPQNQLVIRGRIILHNGAYS-LTLQGLGISWGLRSLRSLRELGSGLALHNN 477
Db 400 QSWPPHMHNFVFSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGRSLKEISAGRITYISAN 459
Qy 478 THLCFVHTVPWDQLFRPHQALLHTA-NRPDECVGEGGLACHOLCARGCHCGPPTQCVN 536
Db 460 RQLCYHHSNLNWKVLRGTERTLDIKHNRPRDCVAEGKVDPLCSCGGCGNGPQGOCLS 519
Qy 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHY 596
Db 520 CRNYSRGVCVTHCNFLNGEPREFAHEAECFCHPECPQMEGTATCNGSGSDTCAQCCHF 579
Qy 597 KDPFCVACRSGVKPDLISYMPIWPKFPDEGACQPCPNCTHSCVDLDDKCPAEQRA-- 654
Db 580 RDGPHCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHENCQCGCKGPELQDCLGQTLVLI 637
Qy 655 --SPLTSIVSAVGLLVLVGLVGVGILI KEROQKIR-KYTMRRLLQTELVEPLTPSGA 711
Db 638 GKTHLTWALTVIAG--LVVIFMLGGTFLYWRGRIQNKRAWRYLERGESIEPLDPS-E 694
Qy 712 MPNAQQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
Db 695 KANKVLARIKPETELKRLKVLGSGVGVGVHGVWIPGESIKIPVICKIVEDKSGRQSFQ 754
Qy 772 EILDEAYMAGVSGPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVHNRGLSGQDLLNW 831
Db 755 AVTDHMLAIGSLDHAHIVRLGLCPGSSLQVLTQVPLGSLLDHVHQRGALGPQLLNLW 814
Qy 832 CMOJAKGMSYLEDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 815 GVQJAKGMYLLEEHGMVHRNLAAARNVLKSPQOVADFVADLLPPDKQLLYSEAKTP 874
Qy 892 IKWMALESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 951
Db 875 IKWMALESIHFGKYTHQSDVMSYGVTVWELMTFGAEPYAGRLAEVPLDLEKGERLAQ 934
Qy 952 ICTTIDVYMWKCMWIDSECRPRPRELVSEFSRWARDQRFVVIQNEIDLGA---SPLDS 1008
Db 935 ICTTIDVYMWKCMWIDENIRPTEKELANEFTRWARDPPRYLVIKRES-GPGIAPGPEPH 993
Qy 1009 TFYRSLLDDDMGDLVDAEEYLVPOQGFPCDPAPAGAGVMVHRHRSSTSGGDLTLG 1068
Db 994 GLTNKULEEVELEPELDLDLDAEED-----NLATTTLSALSPL 1034
Qy 1069 LEP-SEEAAPRSPLAPSGAGSDVFDGLGMAAGLQSLPHTD-PSPLQRYSDPTVPL 1126
Db 1035 VGTLLNRPRGSQSLSPSSGY-MPMNQGNLGSQESAVSGSSERCPRPVSLH-----PM 1087
Qy 1127 P-----SETDGYA-----PLTCSPOPE-----YVNPQDVRPQPPSPRE 1160
Db 1088 PRGCLASESGHVTGSEAELOEKVSMCRSRSRSPRPGDSAYHSQRHSLLTPVTPLS 1147
Qy 1161 GP-----LPAARPAATLERAKTLLSP-GKNGV-----KDVFAFGGAVENPEY 1202
Db 1148 PPGLEEDVNGVMPDTHKTKTPSRECTLSSVGLSSVLGTEEDED-----EVEY 1199
Qy 1203 LTPCGGAAPQHPPPAFSPAFDNLYYMD-----QDPPERGAPPSTFKGTPTAE 1250
Db 1200 MNRRRRHSP-PHPPRPSLEELGYEYMDVGSDLSASLSGSTQCPHVPVIMPTAGTTTPE 1258
Qy 1251 NPEYL 1255
Db 1259 DYEYM 1263
```

RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein

C:Species: Rattus norvegicus (Norway rat)

C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998

C:Accession: JC4387

R:Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protei

A:Reference number: JC4387; MUID:96096535; PMID:8522190

A:Accession: JC4387

A:Molecule type: mRNA

A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:g915309; PID:g915390

A:Experimental source: liver

A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res

C:Comment: This protein is a functional heregulin receptor that transduces signals t

C:Genetics:

A:Gene: ErbB3

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hor

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>

F:640-659/Domain: transmembrane #status predicted <TMM>

F:703-970/Domain: protein kinase homology <KIN>

F:713-721/Region: protein kinase ATP-binding motif

F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 32.9%; Score 2248.5; DB 2; Length 1339;

Best Local Similarity 39.8%; Pred. No. 3.5e-89;

Matches 516; Conservative 168; Mismatches 436; Indels 177; Gaps 36;

```
Qy 3 LAALCRWGLLIALPPGAA---STQVCTGTOMKRLRLPASPEHLDMRLHLYOGCOVQGN 59
Db 7 LQVLC---FLLSLARGSEMNSQAVCPGLTNGLSVTGDADNOVQTLTKLKEKCEVWGN 62
Qy 60 LELTYLPTNASLSFLQDIQEVQGYLYIAHNOVRQVPLORLIRVIRGTQLFEDNYALAVLDN 119
Db 63 LEIVLTGHNDLSFLQWIREVTAYVLVAMNEFSLPLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDPLNNTPVTGASPGGLRELOLRSLTEILLKGVVLIQBNPOLCYQDTILMKDIFHNQQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR--- 170
Qy 180 ALTLLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGG-ARCKGPLPNNFTV 238
Db 171 GABIVVKNNGANGCPCHEVCKG-KCWGPGPDDCQILTKTICAPQCNGRCFGNP----- 223
Qy 239 SFWLVRVPKVSASHLE-----KHSQCLACLHFNHSGICELHCPALVYNTDTPESHNP 290
Db 224 -----NQCCHDEACGCGSPQDTCFACRHFNDGACVPRCPQLVYNKLTFLQEPN 275
Qy 291 PEGRYTFGASCVTACPNYLSLTDVSGSCTLVCPHUNQEVTAEDGTQRCCKSPCARVCVG 350
Db 276 PHTKIQGVGVASCPHNFV-VDTQFCVRACPPDRKMEVD-KHGLKNCEPCGGLCPKACBG 333
Qy 351 LGMEHLREVRVTSANIOBFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPFETLE 410
Db 334 TGSQ--SRYQTVDSNIDGFVNCTKILGNLDFLTGLNVDPMHKIPALDPEKLNVPRTV 391
Qy 411 EITGYLYISAWPDSLPLSVPQNQLVIRGRIILHNGAYS-LTLQGLGISWGLRSLRSLRE 469
Db 392 EITGYLNIQSWPPHMHNFVFSNLTTIGRSLYNRGFSLLIMKNLNVTSLSGRSLKEISA 451
Qy 470 GLALIHNTHLCHFVHTVPWDQLFRPHQALLHTA-NRPDECVGEGGLACHOLCARGCH 528
Db 452 GRVYISANQQLCYCHHSNLNWKVLRGTERTLDIKHNRPRDCVAEGKVDPLCSCGG 511
Qy 529 PGPTQCVCNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCF 588
```


Db 512 PAPGQCLSCRNYSREGVCVTHCNFLOQEPREFVHEAQCSCHPECLPMEGSTSYNGSGD 571
QY 589 QCVACAHYKDPFCVACRPSGVKPDLSYMPWKPFDEGACQPCPINCTHSC--VDLDDK 646
Db 572 ACARCAHFRDGHCVNCPHGLG--AKGPIYKYPDAQNECRCHENCCTQCGNPELQDC 629
QY 647 GCPAEQASPLTSIVSAVVGILLVVLGVVFGILIKRROOKIR-KYTMRLLOETELVER 705
Db 630 LQAEVLMKSPHLVIAVTVG--LAVILMLGGSFYWRGRRIONKRAMRYLBERGESIEP 687
QY 706 LTPSGAMPNQAOQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVURENT 765
Db 688 LDPS-EKANVKLARIFKETELRKLKVLGSGVGTGKHGIWIPGESIKIPVCIKVIEDKS 746
QY 766 SPKANKEIIDEAVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRRGLS 825
Db 747 GRQSOAVTDHMLAVGLDHAHIVRLGLCPGSSQLQVLPGLSLLDHVKQHRETLGP 806
QY 826 QDLLNMCMOIAKMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHA 885
Db 807 QULLNMGVQIAKGMYYLESHMVRDLAARNVLMKSPQVQVADFGVADLLPDDKQLLH 866
QY 886 DGGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLEKGE 945
Db 867 SEAKTPIKWMALLESIFHGKYTHQSDVMSYGVTVWELMTFCAPYAGLRALAEIPDLLEKGE 926
QY 946 RLPOPPICITDVMIMVKCMIDSECRPRFRELVSFSRMDRPORFVVIQNEDLGPASP 1005
Db 927 RLAQPOICITDVMIMVKCMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP 985
QY 1006 LOSTFVRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGVVHRRSSSTRSGGDL 1065
Db 986 --PAEAPSVLTTEL-----QEALEPEL-----DL 1009
QY 1066 TLGLEPSEE-----EAPRSPLAPSEG-----AGSDVFDGDLG 1097
Db 1010 DLDLEAEELGATSLGALSPLTGTLTRPGSQSLSPSSGYMPMQSSIGEACLDSAVL 1069
QY 1098 MGAAGLQSLPHDPSLPORYSEDPVLPSETDGVV-----APL-----TC-----SP 1141
Db 1070 GGREQFSRPLSLH-PIPRGR-----PASESEGHVGTSEALQEKVSVCSRSRSRSP 1121
QY 1142 QPE---VYNQPDVRPQPSPREGP-----LPAARPAGATLERAKTSLP-CXNG 1185
Db 1122 RPRGSAYHSQRHSLTPTPLSPGLCEEDNGVYMPDTHLRGASSSREGTLSSVGLSS 1181
QY 1186 VV-----KDVFAFGAVENPEYLTPOGGAAPQPHPP 1216
Db 1182 VLGTEEEDE-----EEYEYMNKRKRGSP-PRPP 1209
RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C:Species: avian leukosis virus, ALV
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C:Accession: B00643; A00643
R:Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: B00643
A:Molecule type: mRNA
A:Residues: 1-698 <NIL>
C:Cross-references: GB:M10066; GB:M13881; NID:g211749; PID:AAA48763.1; PID:g211750
A>Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C:Comment: This protein is synthesized as a gag-env-erbB protein.
C:Genetics:
A:Gene: gag-env-erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F:194-459/Domains: protein kinase homology <KIN>
F:202-210/Region: protein kinase ATP-binding motif
F:229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 8.1e-69;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 584 GPEADOCVACAHYKDPFCVACRPSGVKPDLSYMPWKPFDEGACQPCPINCTHSCVDL 643
Db 60 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VKYNADANAVCOLCHPNCRTCKGKP 116
QY 644 DDKGCPAEQASPLTSIVSAVVGILLVVLGVVFGILIKRROOKIRKYTMRLLOETEL 702
Db 117 GLEGCP--NGSKTSPISAGVVGGLLCLVVGIGLGLRR-HIVKRTLRLLQOREL 172
QY 703 VPELTPSGAMPNQAOQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLR 762
Db 173 VEPLTPSGEAPNQAHILKETEFKKVKVLGSGAFGTVYKGLWIPEGEKVIPVAIKELR 232
QY 763 ENTSPKANKEIIDEAVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRR 822
Db 233 EATSPKANKEIIDEAVMASVDNPHVCRLLGICLTSTVOLVTOLMPYGCLLDVIREHKN 292
QY 823 LGSQDLLNMCMOIAKMSYLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 293 IGSQYLLNMCVQIAKGMYYLESHMVRDLAARNVLMKSPQVQVADFGVADLLPDDKGADEKE 352
QY 883 YHADGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIPIAREIPDLLE 942
Db 353 YHAEGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFCGPKYDGIPIASEISSVLE 412
QY 943 KGERLPOPPICITDVMIMVKCMIDSECRPRFRELVSFSRMDRPORFVVIQNEDLG 1001
Db 413 KGERLPOPPICITDVMIMVKCMIDADSRPKFRELIAEFSKWARDPPRYLVITQDGRMH 472
QY 1002 PASPLDSTFYRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGVVHRRSSSTRSG 1061
Db 473 LPSPTDSKYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 513
QY 1062 GGDLTGLLEPSEEEAPRSPL-----APSEAGSDVFDGDLGMAAGLQSLPHDPSPLQ 1116
Db 514 -----SRTPLSSLSATSNNSATNCID-----RNGQGHVREDSEFVQ 550
QY 1117 RYSEDPVPLPSET--DCYVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVYVQ--LMPKKPS----- 585
QY 1175 RAKTLSPGKNGVVKVDF-----AFGGAIVENPEYLTPOGGAAPQPHPPAF 1219
Db 586 ----TAMVQNOIYNNISLTAISKLPMDSRYQNSHSTAVDNPEYL-----NTNOSPLA 633
QY 1220 SPAPNLVYWDQ-----DPPE-----RGAPSTFKGTPAENPEYLGIDVP 1260
Db 634 KTVFESSPYWIQSGNHQINLDNPDYQDDFLPNETKENGLLKYPAAENPEYLRVAAP 689
RESULT 12
TVVUHV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C:Species: avian erythroblastosis virus
C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C:Accession: A00644; A38022
R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fa
A:Reference number: A00644; MUID:84026539; PMID:6313229
A:Accession: A00644
A:Molecule type: DNA
A:Residues: 1-604 <YAM>
A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus
C/Species: avian erythroblastosis virus
C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C/Accession: S00727
R/Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M. J.

Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; PMID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.7e-63;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 584 GPEADQVACAHYKDPFCVACPSGVKPDLSYMPIKFPDEEGACQPCPINCTHSCVDL 643
DB 1 GP-DHCKMCAFIIDGPHCVKACPAVLGENDTL-VKRYADANAVCQLCHPNCTRGCKGP 57
QY 644 DDGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 702
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVVGLGIGLYLRR-HIVKRTLRLLQEREL 113
QY 703 VEPLTPSGAMPNOAQRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLR 762
DB 114 VEPLTPSGAPNOAHLRIKETEFKVKVGLGAGFTVYKGLWIPEGEKVTIPVAIKELR 173
QY 763 ENTSPKANKEILDEAVMAGVGSVYSRLGLCLSTVQLTOLMPYGCCLLDHVRNCR 822
DB 174 EATSPKANKEILDEAVMVASVNDPHVCRLLGLCLSTVQLTOLMPYGCCLLDYIREHKN 233
QY 823 LGSODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNWCVOIAKGMNLEERHLVHRDLAARNVLKTPQDKITDFGLAKQLGADEKE 293
QY 883 YHADGCKVPIKMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 942
DB 294 YHAEQCKVPIKMALESILHRIYHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353
QY 943 KGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPPICITIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLG 413
QY 1002 PASPLDSTFYRSLLEDDDMGLVDAAEYLVPOQGFCDPAPGAGMVHRRHSSSTRSG 1061
DB 414 LPSPTDSKXYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1062 GGDLTGLLEPSEEAAPRPL-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQ 1116
DB 455 -----SRTPLSSLSATSNNSATNCIDRNG-----H----- 481
QY 1117 RYSEDPVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGAT-LER 1175
DB 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVILTAISK 523
QY 1176 AKTLSPGKNGVVKVFAFGGAVENPEYL 1203
DB 524 LPMDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
B44776
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
C:Species: avian erythroblastosis virus
C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C:Accession: B44776
R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A:Title: Six amino acids from the retroviral gene gag greatly enhance the transforming P
A:Reference number: A44776; PMID:90206603; PMID:1969616

A:Accession: B44776
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BRU>
A:Cross-references: GB:X52211
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif

Query Match 23.8%; Score 1623; DB 2; Length 540;
Best Local Similarity 54.9%; Pred. No. 8.8e-63;
Matches 340; Conservative 69; Mismatches 120; Indels 90; Gaps 14;

QY 593 CAHYKDPFCVACPSGVKPDLSYMPIKFPDEEGACQPCPINCTHSCVDLDDKGPAPQ 652
DB 3 CAHFIDGPHCVKACPAVLGENDTL-VKRYADANAVCQLCHPNCTRGCKPGLEGCP--- 58
QY 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGA 711
DB 59 NGSKTPSIAAGVVGGLCLVVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGE 117
QY 712 MPNOAMRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
DB 118 APNOAHLRIKETEFKVKVGLGAGFTVYKGLWIPEGEKVTIPVAIKELREATSPKANK 177
QY 772 EILDEAVMAGVGSVYSRLGLCLSTVQLTOLMPYGCCLLDHVRNCRGLSQDLLNW 831
DB 178 EILDEAVMVASVNDPHVCRLLGLCLSTVQLTOLMPYGCCLLDYIREHKNIGSQYLLNW 237
QY 832 CMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGCKVP 891
DB 238 CVQIAKGMNLEERHLVHRDLAARNVLKTPQDKITDFGLAKQLGADEKEYHAEQCKVP 297
QY 892 IKMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 951
DB 298 IKMALESILHRIYHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
QY 952 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
DB 358 ICTIDVYIMVKCMWIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGPASPLDSTF 417
QY 1011 YRSLLEDDDMGLVDAAEYLVPOQGFCDPAPGAGMVHRRHSSSTRSGGDLTLGLE 1070
DB 418 YRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPRPL-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVTP 1125
DB 450 -----SRTPLSSLSATSNNSATNCIDRNG-----H----- 476
QY 1126 LPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGAT-LERAKTLSPGKN 1184
DB 477 -PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVILTAISKLPMDSRYN 527
QY 1185 GVVKDVFAGGAVENPEYL 1203
DB 528 -----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:08:59
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-210-224-14
Perfect score: 6827
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6684	97.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5872	86.0	1257	1 ERB2_RAT	P06494 rattus norv
3	5870.5	86.0	1254	1 ERB2_MESAU	Q05053 mesocricetu
4	3052	44.7	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3031	44.4	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2905.5	42.6	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2886	42.3	1308	1 ERB4_RAT	Q62956 rattus norv
8	2618.5	38.4	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2342.5	34.3	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2271.5	33.3	1339	1 ERB3_RAT	Q62799 rattus norv
11	1883	27.6	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.6	634	1 ERBB_ALV	P00534 avian leuko
13	1703	24.9	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1510	22.1	703	1 EGFR_CHICK	P13387 gallus gall
16	1351	18.3	1323	1 LTR3_CABEL	P24348 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70424 mus musculu
18	678	9.9	1382	1 INSR_HUMAN	P06213 homo sapien
19	671	9.8	1363	1 ILPR_BRALA	Q02466 brachiosto
20	671	9.8	1383	1 INSR_RAT	P15127 rattus norv
21	670.5	9.8	1372	1 INSR_MOUSE	P15127 rattus norv
22	662	9.7	1297	1 IRR_HUMAN	P14616 mus musculu
23	661	9.7	1607	1 MIPR_LYMST	P14616 mus sapien
24	659	9.7	1300	1 IRR_MOUSE	Q25410 lymnaea sta
25	656.5	9.6	1300	1 IRR_CAVPO	Q9wt14 mus musculu
26	643	9.4	1477	1 HTK7_HYDAT	P14617 cavia porce
27	604	8.8	1367	1 IGR1_HUMAN	Q25197 hydra atten
28	597	8.7	1373	1 IGR1_MOUSE	P08069 homo sapien
29	593.5	8.7	1370	1 IGR1_RAT	Q60751 mus musculu
30	591	8.7	987	1 EPB4_HUMAN	P24062 rattus norv
31	589.5	8.6	977	1 EPB2_MOUSE	Q03145 mus sapien
32	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien
33	588	8.6	1390	1 INSR_AEDAE	Q93105 aedes aegyp

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell			
DE	surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR NGL OR NEU.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erbB-2 gene to			
RT	epidermal growth factor receptor."			
RL	Nature 319:230-234(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Frankie U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene."			
RL	Science 230:1132-1139(1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma."			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization."			
RL	Genomics 15:426-429(1993).			
CC	!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	!- SUBCELLULAR LOCATION: Type I membrane protein.			

ALIGNMENTS

34	584.5	8.6	976	1	EPB2_HUMAN
35	583.5	8.5	984	1	EPB1_CHICK
36	581	8.5	987	1	EPB4_MOUSE
37	581	8.5	2146	1	INSR_DROME
38	573.5	8.4	1053	1	PAK1_CHICK
39	572.5	8.4	984	1	PAK1_RAT
40	569	8.3	1088	1	PAK1_XENLA
41	567	8.3	757	1	HT16_HYDAT
42	566.5	8.3	984	1	EPB1_HUMAN
43	564.5	8.3	985	1	EPBA_XENLA
44	564	8.3	902	1	EPBB_XENLA
45	563	8.2	1052	1	PAK1_MOUSE

P29317	homo sapien
Q07494	gallus gall
P54761	mus musculu
P09208	drosophila
Q00944	gallus gall
P09759	rattus norv
Q91738	xenopus lae
P53356	hydra atten
P54762	homo sapien
Q91571	homo sapien
Q91736	xenopus lae
P34152	mus musculu


```
QY 713 PNOAQRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 766
QY 773 ILDEAYVMAGVGSPPYVRLGICLCTSTVQLVLTOMPVYGCCLLDHVRENRGLSGSODLLNWC 832
Db 767 ILDEAYVMAGVGSPPYVRLGICLCTSTVQLVLTOMPVYGCCLLDHVRENRGLSGSODLLNWC 826
QY 833 MQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDDIDETEHADGKVPPI 892
Db 827 MQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDDIDETEHADGKVPPI 886
QY 893 KWALESILRRRTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIDPLEKGERLPQPI 952
Db 887 KWALESILRRRTHOSDVMSYGVTVWELMTFCAKPYDGIPIAREIDPLEKGERLPQPI 946
QY 953 CTIDVTMIMVKMWIDSECRPRELVSFSESRMAROPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVTMIMVKMWIDSECRPRELVSFSESRMAROPQRFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SLEDDDDMGDLVDAEYLYPQQGFCFDPDPAGGVMVHRRHSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLYPQQGFCFDPDPAGGVMVHRRHSSSTRSGGDLTLGLEPS 1066
QY 1073 EEAAPRSPLAPSGAGSDVPDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1132
Db 1067 EEAAPRSPLAPSGAGSDVPDGLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLPCKNGVVKDVFA 1192
Db 1127 YVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLPCKNGVVKDVFA 1186
QY 1193 FCGAVENPEYLTQGGAAPOHPPPAFSPAFDNLXYWDQDPPPERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTQGGAAPOHPPPAFSPAFDNLXYWDQDPPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255

RESULT 2
ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 5:691-704(1991).
```

```
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSSP; P11362; LFQK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 339 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```

FT DISULFID 517 530 BY SIMILARITY.
FT FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 6129264583011402 CRC64;
Query Match 86.0%; Score 5872; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 3.7e-311;
Matches 1090; Conservative 51; Mismatches 106; Indels 24; Gaps 4;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMLRHLRYGCGQVQGNL 60
Db 1 MELAAMCRWGLLLALLPPGIAGTCTGCTDMKRLPASPETHLDMLRHLRYGCGQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDTQEVGYVLIAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYVPANASLSFLQDTQEVGYVLIAHNOVKRVLQRLRIVRGTOLEFEDNYALAVLDNR 120
Qy 121 DPLANTTPVT-GASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYODTILKWDIFHKNNOL 179
Db 121 DPQNVAASTPRTPEGLRELQRLSLTEILKGGVLIQIRNPOLCYQDWLWKDKVFRKNNOL 180
Qy 180 ALTLLIDNRSRACHPCSPMKGSRGWESSEDCSLTRTVGAGCCKPGLPENNFTVS 239
Db 181 APVDIDNRSRACHPCCAPACKDNHGWESPEDCQILTGCTGSCARCKGLP----- 233
Qy 240 FWLVPKVSASHLE-----KHSDCGLAHFNHSGICELHCPALVTYNTDTFESMNP 291
Db 234 -----TDCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMNP 286
Qy 292 EGRYTFGASCVTCPYNYLSDVGSCTLVCPHNLQEVTAEDGTORCEKCKSPCARVCYGL 351
Db 287 EGRYTFGASCVTCPYNYLSTEVGSCTLVCPNNQEVTAEDGTORCEKCKSPCARVCYGL 346
Qy 352 GMEHLRVRVTSANIOEFAGCKKIFGSLAFPLSFDFGDPASNTAPLOEQLOVFETLEE 411
Db 347 GMEHLRGARAITSDNVQEFDFGCKKIFGSLAFPLSFDFGDPSSGIAPLRPELOQLOVFETLEE 406
Qy 412 ITGYLYISAWPDSLPLDLSVFONLOVIRGRILLHNGAYSITLQGLGISMGLRLSRELQSL 471
Db 407 ITGYLYISAWPDSLURDLSVFONLRIIRGRILLHNGAYSITLQGLGISHLSRELQSL 466
Qy 472 ALIHNTLHLCFVHTVPMDQLFRNPHQALLHTANPEDE-CVGEGLACHQLCARGHCWGP 530
Db 467 ALIHNAHLCFVHTVPMDQLFRNPHQALLHSGNPEEDLVCYSSGLVCNSLCAHCHGWGP 526
Qy 531 PTQCVNCSQFLRGQECVECRVLQGLPREYVNAHCLPCHPCEQCPQNGSVTCFQPEADQC 590
Db 527 PTQCVNCSHFLRGQECVECRVWKGKLPREYVSDKRLCPCHPEQCPQNSSETCFQSEADQC 586
Qy 591 VACHAYKDPFCVACRSPGVKPDLSYMPIWKFPEEGACQPCPNCNTHSCVDLDDKCPA 650
Db 587 AACAHYKDSKSSVACRSPGVKPDLSYMPIWKFPEEGICQPCPNCNTHSCVDLDERGCPA 646
Qy 651 EQRASPLTSIVSAVVGILLVVLGVWFGLIKRQKIRKYTMRLLOETELVELPLTPSG 710
Db 647 EQRASPTFIATVVGILLVVLGVWFGLIKRQKIRKYTMRLLOETELVELPLTPSG 706
Qy 711 AMPNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKAN 770

Db 707 AMPNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKAN 766
Qy 771 KEILDEAYVWAGVSPYVSRLLGLCLTSTVQLVTLQMLPYGCLLDHVRNRRGLSGQDLN 830
Db 767 KEILDEAYVWAGVSPYVSRLLGLCLTSTVQLVTLQMLPYGCLLDHVRHRRGLSGQDLN 826
Qy 831 WCMQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKV 890
Db 827 WCVQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKV 886
Qy 891 PIKMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 950
Db 887 PIKMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLP 946
Qy 951 PICTIDYMIWVKCWMIDSECRPRFRELVSFESFMRDPPQRFVIVIONEDLGPASLDSTF 1010
Db 947 PICTIDYMIWVKCWMIDSECRPRFRELVSFESFMRDPPQRFVIVIONEDLGPASLDSTF 1006
Qy 1011 YRSLEDDMDGLVDAAEYLVPOQFFCPDAPCAGGMVHRRSSSTRSCGGDLTLGLE 1070
Db 1007 YRSLEDDMDGLVDAAEYLVPOQFFCPDAPCAGGMVHRRSSSTRSCGGDLTLGLE 1066
Qy 1071 PSEBAPRSLAPSEGAGSDVFDGLQMGAAKGLQSLPTHDPSPQLRYSDEPTVPLPSET 1130
Db 1067 PSEBAPRSLAPSEGAGSDVFDGLQMGAAKGLQSLPTHDPSPQLRYSDEPTVPLPSET 1126
Qy 1131 DGIVAPLTCSPQPEYVNVQDVRPOPSPRGPPLPAAPAGATLERAKTSLSPGKGVWVDV 1190
Db 1127 DGIVAPLTCSPQPEYVNVQDVRPOPSPRGPPLPAAPAGATLERAKTSLSPGKGVWVDV 1186
Qy 1191 FAFGAVENPEYLTPOCGAAPPHPPPAFPAFONLYWDDPPERGAAPPSTFKCTPTAE 1250
Db 1187 FAFGAVENPEYLVPRGTASPPHSPAFPAFONLYWDDPPERGAAPPSTFKCTPTAE 1246
Qy 1251 NPEYLGLDVVPV 1261
Db 1247 NPEYLGLDVVPV 1257
RESULT 3
ID ERB2 MESAU STANDARD; PRN; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

```
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D16295; BAA03801.1; -.
CC HSSP: P11362; 1FGK
CC InterPro: IPR000494; EGFR_L domain.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002174; Furin-like.
CC InterPro: IPR001245; Tyr_pkinase.
CC InterPro: IPR004019; YLP motif.
CC Pfam: PF00069; pkinase; 1.
CC Pfam: PF00757; Furin-like; 1.
CC Pfam: PF01030; Recep_L domain; 2.
CC Pfam: PF02757; YLP; 2.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00261; FU; 3.
CC SMART: SM00219; TyrKc; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
CC DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 653 675 POTENTIAL.
CC DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 158 368 CYS-RICH.
CC DOMAIN 472 644 CYS-RICH.
CC DOMAIN 720 987 PROTEIN KINASE.
CC NP_BIND 726 734 ATP (BY SIMILARITY).
CC BINDING 753 753 ATP (BY SIMILARITY).
CC ACT_SITE 845 845 BY SIMILARITY.
CC DISULFID 195 204 BY SIMILARITY.
CC DISULFID 199 212 BY SIMILARITY.
CC DISULFID 236 244 BY SIMILARITY.
CC DISULFID 240 252 BY SIMILARITY.
CC DISULFID 255 264 BY SIMILARITY.
CC DISULFID 268 295 BY SIMILARITY.
CC DISULFID 299 311 BY SIMILARITY.
CC DISULFID 315 331 BY SIMILARITY.
CC DISULFID 334 338 BY SIMILARITY.
CC DISULFID 511 520 BY SIMILARITY.
CC DISULFID 515 528 BY SIMILARITY.
CC DISULFID 531 540 BY SIMILARITY.
CC DISULFID 544 560 BY SIMILARITY.
CC DISULFID 563 576 BY SIMILARITY.
CC DISULFID 567 584 BY SIMILARITY.
CC DISULFID 587 596 BY SIMILARITY.
CC DISULFID 600 623 BY SIMILARITY.
CC DISULFID 626 634 BY SIMILARITY.
CC DISULFID 630 642 BY SIMILARITY.
CC MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
CC VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
CC VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;
```

```
Query Match 86.0%; Score 5870.5; DB 1; Length 1254;
Best Local Similarity 84.9%; Pred. No. 4.5e-311;
Matches 1088; Conservative 57; Mismatches 89; Indels 47; Gaps 4;

Qy 1 MELAAALCRWGLLLALLPPGAASSTOVCTGTDMLKRLPASBETHLDMLSHLYOGCVOVGNL 60
Db 1 MELAAACGMLLLALLSPGASGTQVCTGTDMLKRLPASBETHLDIVRHLYOGCVOVGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIARNOVQVPLQRLRIVRGTQLPEDNYALAVLDNG 120
Db 61 ELTYLPANATLSFLQDIQEVQGYVLIARNOVQVPLQRLRIVRGTQLPEDNYALAVLDNR 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHNQOLA 180
Db 121 DPLDNVTTATRTPEGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHNQOLA 180
Qy 181 LTLIDTNRSPACHPCSPCKSGSCWGESSEDCSLRTV-----C 220
Db 181 PVDIDTNRSPACHPCSPCKSGSCWGESSEDCSLRTV-----C 220
Qy 221 AGGCARCKGPLPFNNFTVSWFLRVKVSASHLEKSHSDCLACLFHNSGICELHCPALVTY 280
Db 221 AGGCARCKGPLPFNNFTVSWFLRVKVSASHLEKSHSDCLACLFHNSGICELHCPALVTY 280
Qy 241 AAGCT---GP-----KXSDCLACLFHNSGICELHCPALVTY 274
Db 241 AAGCT---GP-----KXSDCLACLFHNSGICELHCPALVTY 274
Qy 281 NTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVSGCTLVCPLNHNSGICELHCPALVTY 340
Db 275 NTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVSGCTLVCPLNHNSGICELHCPALVTY 334
Qy 341 SKPCARVCYGLGMEHLREVRANTSAIOBFGCKKIFGSLAFIPESDGPASNTAPLOP 400
Db 335 SKSCARVCYGLGMEHLRGARAITSAIOBFGCKKIFGSLAFIPESDGPASNTAPLOP 394
Qy 401 EQLOVFETLEETGYLIYSAWPDLSVFNQVIRGRILHNGAYSLSLQGLIGISWLG 460
Db 395 EQLOVFETLEETGYLIYSAWPDLSVFNQVIRGRILHNGAYSLSLQGLIGISWLG 454
Qy 461 LRSRLRELGSLALIHNTLHLCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLACHOL 520
Db 455 LRSRLRELGSLVLIHRNTHLCFVHTVPDQLFRNPHQALLHSGNPSEECGLKDFACVPL 514
Qy 521 CARGHCWGPGTQCVNCSQFLRQECVCECRVLQGLPREYVNAHCLPCHPECOQNGSV 580
Db 515 CARGHCWGPGTQCVNCSHFLRQECVCECRVLQGLPREYVNAHCLPCHPECOQNGSV 574
Qy 581 TCFGPEADOCVACAHYKDPFPCVARGCPGVKPDLSYMPIWKFDEEGACOPCPINCTHSC 640
Db 575 TCTGSEADOCVACAHYKDPFPCVARGCPGVKPDLSYMPIWKFDEEGACOPCPINCTHSC 634
Qy 641 VLDDKGCPCAEQASPLTSIVSAVVGILLVVVGVVFGILLIKRRQOKIRKYTMRRLLQET 700
Db 635 VLDDKGCPCAEQASPLTSIVSAVVGILLVVVGVVFGILLIKRRQOKIRKYTMRRLLQET 694
Qy 701 ELVEPLTPSGAMPNQAMRILKETELRKVKVGLSGAFGTVYGIWIIPDGENVKIPVAIKV 760
Db 695 ELVEPLTPSGAMPNQAMRILKETELRKVKVGLSGAFGTVYGIWIIPDGENVKIPVAIKV 754
Qy 761 LRENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTLQMPYCLLDHVRNR 820
Db 755 LRENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVQLVTLQMPYCLLDHVRNR 814
Qy 821 GRIGSQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 880
Db 815 GRIGSQDLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 874
Qy 881 TEYHADGGKVPKWMALSIILRRRFTHQSDVWSYGYVTWELMTFGAKPYDGIIPAREIPDL 940
Db 875 TEYHADGGKVPKWMALSIILRRRFTHQSDVWSYGYVTWELMTFGAKPYDGIIPAREIPDL 934
Qy 941 LEKGERLPOPPICITIDVYIMVWKCMIDSECPFRFRELVSFSESRMARDPORFVVIQNEDL 1000
Db 935 LEKGERLPOPPICITIDVYIMVWKCMIDSECPFRFRELVSFSESRMARDPORFVVIQNEDL 994
```



```
Qy 957 VYIMVKKWMIDSECRPRFRELVSFGRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSL 1015
Db 943 VYIMVKKWMIDADSRKRELIIEFSKMDRDPQRYLVICQDERMHLPSPTDSFYRALM 1002
Qy 1016 EDDMDGLDVAEEVLPQCGFFCDPPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEE 1075
Db 1003 DEEDMDVDVADEYLIPOQGGF-----SSPS 1028
Qy 1076 APRSPLAPSGAGSDVFDGLGMAAGKLOSLTHDPSPQLQRYSEDPTVLPSET--DGY 1133
Db 1029 TSRTPLLSLSATSN--NSIVACIDRNLGLOSCPIKEDSFQRYSSDPTGALTEDSIDT 1086
Qy 1134 VAPLTCSPQEVYNQDVRPQPPSPREGPLPAARAGATLAKTLSPGKNGVVKQVFAF 1193
Db 1087 L-----PVPEYINQ-SVPKRPAGSVQNPVYHNP-----APSRDPHYQD--PH 1129
Qy 1194 GGAIVENPEYL-TPOGGAAPQHPPPAPSPAFDNLNYYDQ-----DP-----PER 1236
Db 1130 STAVGNPEYLTQV-----PTCVNSTFDSPAHWAQKGSQHSQISLDNPDYQODFFPKE 1180
Qy 1237 GAPSPTEFGTPTAENPEYL 1255
Db 1181 AKPNGIFKGS-TAENAYL 1198
RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX Hibbs M.L.;
RA Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413 (1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91212866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
```

```
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676 (1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eslinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA55587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSP; P11362; IFGK
CC MGD; MG1:95294; Egfr.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL
CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
CC DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 648 670 POTENTIAL.
CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
CC REPEAT 75 300 APPROXIMATE.
CC REPEAT 390 600 APPROXIMATE.
CC DOMAIN 1028 1071 SER-RICH.
CC DOMAIN 714 981 PROTEIN KINASE.
CC NP_BIND 720 728 ATP (BY SIMILARITY).
CC BINDING 747 747 ATP (BY SIMILARITY).
CC ACT_SITE 839 839 BY SIMILARITY.
CC DISULFID 190 199 BY SIMILARITY.
CC DISULFID 194 207 BY SIMILARITY.
CC DISULFID 215 223 BY SIMILARITY.
CC DISULFID 219 231 BY SIMILARITY.
CC DISULFID 232 240 BY SIMILARITY.
CC DISULFID 236 248 BY SIMILARITY.
CC DISULFID 251 260 BY SIMILARITY.
CC DISULFID 264 291 BY SIMILARITY.
CC DISULFID 295 307 BY SIMILARITY.
CC DISULFID 311 326 BY SIMILARITY.
CC DISULFID 329 333 BY SIMILARITY.
```


epidermal growth factor receptor family.";
Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
(2)
SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
TISSUE=Fetal brain;
MEDLINE=97476287; PubMed=9334263;
Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
Klagbrun M.;
"A novel juxtamembrane domain isoform of HER4/ErbB4. Isoform-specific
tissue distribution and differential processing in response to
phorbol ester";
J. Biol. Chem. 272:26761-26768(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
CC PITUITARY, SPLEEN, TESTIS AND PANCREAS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L07868; AAB59446.1; -;
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3432; ERBB4.
DR MIM; 600543;
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP motif.
DR Pfam; PF0069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 718 985
PROTEIN KINASE.

FT NP BIND	724	732	ATP (BY SIMILARITY).
FT BINDING	751	751	ATP (BY SIMILARITY).
FT ACT_SITE	843	843	BY SIMILARITY.
FT DISULFID	189	197	BY SIMILARITY.
FT DISULFID	193	205	BY SIMILARITY.
FT DISULFID	213	221	BY SIMILARITY.
FT DISULFID	217	229	BY SIMILARITY.
FT DISULFID	230	238	BY SIMILARITY.
FT DISULFID	234	246	BY SIMILARITY.
FT DISULFID	249	258	BY SIMILARITY.
FT DISULFID	262	289	BY SIMILARITY.
FT DISULFID	293	304	BY SIMILARITY.
FT DISULFID	308	323	BY SIMILARITY.
FT DISULFID	326	330	BY SIMILARITY.
FT DISULFID	503	512	BY SIMILARITY.
FT DISULFID	507	520	BY SIMILARITY.
FT DISULFID	523	532	BY SIMILARITY.
FT DISULFID	536	552	BY SIMILARITY.
FT DISULFID	555	569	BY SIMILARITY.
FT DISULFID	559	577	BY SIMILARITY.
FT DISULFID	580	589	BY SIMILARITY.
FT DISULFID	593	614	BY SIMILARITY.
FT DISULFID	617	625	BY SIMILARITY.
FT DISULFID	621	633	BY SIMILARITY.
FT MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	181	181	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	233	233	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC	626	648	NGPTSHDCIYPTWGTGHTSLPQHA -> IGSSIEDCIGLMD (IN ISOFORM JM-B).
FT SEQUENCE	1308 AA;	146807 MW;	SE4E80985D88761 CRC64;
Query Match	42.6%;	Score 2905.5;	DB 1; Length 1308;
Best Local Similarity	44.7%;	Pred. No. 3.1e-150;	
Matches 608;	Conservative 181;	Mismatches 376;	Indels 195; Gaps 31;
QY 9	WGLLLALLPPGAA	-----STQVCTGTDMLRLPASBETHLDMRLHLVQCGVQVQGNLELTY 64	
Db	8 WWSLLVAAAGTVQPSDSQSVCACTENKLSSLDLEQQYRALRKYYENCEVWNGNLEITS 67		
QY 65	LPTNASLSFLQDIOEVQGVYLIHNRQVPLQRIRIVRGTOQLFEDNYALAVLDNGDPLN 124		
Db 68	IEHNRDLSFLRSRVREVTGVLVALNQFVLPLENLRIINGTKLYEDRYALAIPLNYRKDG 127		
QY 125	NTPPTVGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILMKDKIFHKNQALTLI 184		
Db 128	NP-----GLOELGLKLNLTILNGGVYVDQNKFLCYADTIHWDIVRNWPNSNLTIV 178		
QY 185	DTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAGGC-ARCKGPLFPNNFTVSWLR 243		
Db 179	STNGSGGCRCHKCTG-RCWGPTENHCOTLTRTCAEODCGRCTGYP-----VS----- 227		
QY 244	VPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFFSMPNPEGRY 295		
Db 228	---DCCHRECAAGCGSPKDTDCPACMFNDSGACVTCQPTQFVFNPTTFQLEHFNAYK 283		
QY 296	TFGASCVTACPNYLSLVGSGCTLVCPHLNHOEVTADGTQRCCEKSKPCARVCYGLGMH 355		
Db 284	TYGAFCKVKCPHNFV-VDSSSCVACRACPSKMEV-EENGKMKCPCTDICKPKADGIGTGS 341		
QY 356	LREVRVNTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTLEITY 415		

Db 342 LMSAQVDSNDIKFNCTKINGNLIFLVTHGHPYNAIEAIDPEKLNVRFTVREYTG 401
Qy 416 LYSAMPDLPDLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLASRLSGALAIH 475
Db 402 LNIQSPPNMTDFSVSNLVTIGRVLYSGLSLLKQOGITSLQFQSLAEISAGNIYIT 461
Qy 476 HNTLFCFVHTVPDQQLFRNPQALLHTANRPEDECVEGGLACHQLCARGCHWGPQCV 535
Db 462 DNSNLCCYHTTNTTIFSTINQIRIVIRDNRAENCTAEGMVCNHLSSDGCWGPDPQCL 521
Qy 536 NCQFQRLGQBCVEECVRLQGLPREYVNAHCLPCHPECOF-ONGSVTCFCEPADQVACA 594
Db 522 SCRRFRGRICIESCNLYDGEFREFNGSICVCDPOCEKVEDGLLTCGPGDNCYKCS 581
Qy 595 HYKDPFCFCVACPSGVKPDLSYMPIMKFPDEEGACOPCPINCHTSHSCVDLDDKGC- 648
Db 582 HFKDGPNCEKCPDGLQANSF--IPKYADPDRECHPCNCTCGNGPTSHDCIYYPT 639
Qy 649 -----PAEQRASPLTISVAV--GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQTEL 702
Db 640 GHSTLPQHAR-TPL--IAAGVIGLFLVTVGLTFAVYVRKSGIK-KKRALRRFL-ETEL 694
Qy 703 VEPLTSGAMPNQAORILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAKVLR 762
Db 695 VEPLTSGTAPNQAORILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAKVLR 754
Qy 763 ENTSPKANKEILDYAVMAGVGGPYVRSRLGICLTSTVQLVTQMLPYGCLLDHVRNGR 822
Db 755 ETTGPKANVEFMDEALIMASMDPHLVRLVGLVCLSPITQLVTQMLPHGCLLEYVHEKDN 814
Qy 823 LGSQDLNLCWQIAGKMSYLEDVRLVHRDLAARNLVKSPNWKITDFGLARLLDIDETE 882
Db 815 IGSQDLNLCWQIAGKMSYLEDVRLVHRDLAARNLVKSPNWKITDFGLARLLDIDETE 874
Qy 883 YHAGGKVPKWMALSIILRRRTHOSDVMSYGVYVWELMTFCAPKPDGIPAREIDPLE 942
Db 875 YHAGGKVPKWMALSIILRRRTHOSDVMSYGVYVWELMTFCAPKPDGIPAREIDPLE 934
Qy 943 KGERLPQPPCTIDVTVMYKWMIDSECRPRELVSEFMRARDPQREWVQIONED-LG 1001
Db 935 KGERLPQPPCTIDVTVMYKWMIDSECRPRELVSEFMRARDPQREWVQIONED-LG 994
Qy 1002 PASPLSTYRSLLEDGDLVDABEYLVQOQFCPDPAAGAGVHHRRSSSTRSG 1061
Db 995 LPSPNDSKFFQNLDELEDMDABEYLVQ-PAFNIPP-----IYTSRARDISNRS- 1046
Qy 1062 GGDLTGLSPSEEAERS-----PLAP-SEGAGSDVF 1092
Db 1047 -----EIGHSPPPAYTPMSGNQFYVRDGGFAAEGSVYPRAPTSTIPEAPVAGATAEIF 1102
Qy 1093 DGDLMGAAGLQSLPHTDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEY 1145
Db 1103 DQSCCNGLRKVPVPHVQEDSSQRYSDPTVAPERSPRGELDEEGYMTMRDKPKQEY 1162
Qy 1146 VNOPVDVPPSPREGPLPAARPAAGATLERAKTSLSPKQGVKVDVFAFGGAVENPEYLT 1205
Db 1163 LNPVE-----ENPFVSRR-----KNGDLQ-----ALDNPEYHNA 1191
Qy 1206 QGGAAPQPHPPA-----FSPAFLNLYWDQ 1231
Db 1192 SNG-----PPKAEDYVNEPLYNTFANTLGAKEYLKNILSMPEKAKKAFONPDYWNH 1245
Qy 1232 DPPERGA--PPSTFKGTPT-----AENPEYL 1255
Db 1246 SLPPRSTLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1285
RESULT 7
ERB4 RAT
ID ERB4 RAT
AC Q62956; Q922N7; PRT; 1308 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAC08899.1; -
CC EMBL; U52531; AAC53051.1; -
CC HSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00059; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.

Db 1191 ASSG-----PPKAEDEYVNEPLYLNTFTVALGNAEYMKNSLLSVPEKAKAFNDPDYWN 1244
 QY 1231 QDPPERGA--PPSTFKGTPT-----AENPEYL 1255
 Db 1245 HSLPPRSTLQHPDYLBQYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8

XRMR_XIPMA STANDARD; PRT; 1167 AA.
 AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XRMR OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueuer W., Raulf F.,
 RA Telling A., Robertson S.M., Scharlt M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus."; Nat 341:415-421(1999).
 RN [2]
 RP REVISION TO 515.
 RA Scharlt M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X16891; CAA34770.2; -;
 CC PIR; S06142; S06142.
 CC HSSP; P11362; 1FGK.
 CC InterPro: IPR000494; EGF L domain.
 CC InterPro: IPR000719; Euk_pkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR022290; Ser_thr_pkinase.
 CC InterPro: IPR001245; Tyr_pkinase.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L domain; 2.
 CC PRINTS; PR00109; TYRKINASE.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 5.
 CC SMART; SM00220; S_TKc; 1.
 CC SMART; SM00219; TyrKc; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167
 FT DOMAIN 710 977
 FT NP_BIND 716 724
 FT BINDING 743 743
 FT ACT_SITE 835 835
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 220 228
 FT DISULFID 224 236
 FT DISULFID 237 245
 FT DISULFID 241 253
 FT DISULFID 256 265
 FT DISULFID 269 296
 FT DISULFID 300 311
 FT DISULFID 315 330
 FT DISULFID 333 337
 FT DISULFID 504 513
 FT DISULFID 508 521
 FT DISULFID 524 533
 FT DISULFID 537 553
 FT DISULFID 556 569
 FT DISULFID 560 577
 FT DISULFID 593 615
 FT DISULFID 618 626
 FT DISULFID 622 634
 FT CARBOHYD 114 114
 FT CARBOHYD 144 144
 FT CARBOHYD 201 201
 FT CARBOHYD 356 356
 FT CARBOHYD 365 365
 FT CARBOHYD 398 398
 FT CARBOHYD 417 417
 FT CARBOHYD 501 501
 FT CARBOHYD 576 576
 FT CARBOHYD 621 621
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.4%; Score 2618.5; DB 1; Length 1167;
 Best Local Similarity 44.5%; Pred. No. 9.9e-135;
 Matches 568; Conservative 164; Mismatches 396; Indels 147; Gaps 29;

QY 4 AALRCWGLLLALLPPGAAST---QVCTGTDMLKRLPASPEHLDMLRLHLYQGCOVQGN 59
 Db 8 AALLQ--LLLVLSIRCCSTDPDRKVCQGTSNQMTM---LDNHVLMKMKWYSGCNVULEN 62
 QY 60 LELTYLPTNASLFLQDIOIEVOGVYLIANOVQVPLQRLIRVGTOLFEDNYALAVLDN 119
 Db 63 LEITYTQENODLSFLQSIQEVGVYLIANNEVSTIPLVNLRLIRGQNLVEGFTLLVMSN 122
 QY 120 GDPINNTPTVTGASPGGLRELQLRLSLTEILKGGVLIQRLNQLCYQDTILWKDIFHKNNQL 179
 Db 123 YQK-NPSSP--DVYQVGLKQLQLSLNLTLEILSGGVKSHNPLLCNVEITINMWDIVDKTSNP 179
 QY 180 ALTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLRTRVCAGCG-ARCKGPLPEN--NF 236
 Db 180 TNNLIIPHAFERQCKQKDHGCVNGSWAPGPGHCKQKFTKLICAEQNCRCRCRCPKPIDCNE 239
 QY 237 TVSFWRVRPKVSASHLEKHSKSLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
 Db 240 HCAGGCTGPRA-----TDCLACRDFNDGDKTCPPPKIYDIVSHVQVVDVFNPKYKT 291
 QY 297 FGASCVTACPNYVLTSDVSGCTLVCPLNQOEVTAEQTCRCEKSKCARVCYGLGMEHL 356
 Db 292 FGACVKECPNSYVVTG--GACVRSKPCDGVCPKVCVDGIGSL 349
 QY 357 REVRAVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQQLQVFETLEEITGYL 416
 Db 350 SNTIAVNSTNIRSFNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLWNLTITVKEITGYL 409
 QY 417 YISAWPDSLPDLGVFQNLQVIRGRIHLNGAYS-LTLQGLGISWLGRLSLRELGSGLALIH 475

```
Db 410 VIMMPEMTSLSFQLEIIRGRTTFSRGFSFVVQVRHLQMLGLSLKESVAGNVILK 469
Qy 476 HNTLCLFVHTVPDOLERNPHOALLHTANRPEDECVEGLACHOLCARGHGWGPPTQCV 535
Db 470 NTQLRLVANTINMRLFRSDQSLEYDART-----ENOTCNNECEDGCGWGPPTMCV 522
Qy 536 NCSOFLRGQCEBECRVQLGLPREYVYVNRHCLPCHPECQONGSVTCFGEADOCVACAH 595
Db 523 SCLHVDGGRGVASCNLLQGEPREAQVDGRVCVQCHQCEQLVQTDLSLTCYVGPANCSKSAH 582
Qy 596 YKDPFFVCARCPGVKPDLSYMPIWKFPDEBACQPCPINCSTHSCVLDLDDKGCBAERAS 655
Db 583 FQDGPQICPRPHGILGDGTL-IMKYADKMGQCQCHONCQCGSGFLSGCRGD-IVS 640
Qy 656 PLTSIVSAVGVILLVVLGVVFGILIKRROOKIRKTYMRLRLLQELTELVPELTPSGAMPNQ 715
Db 641 HSSLAVGLVGLLITVIVALLIVVLLRRRIK-RKRIIRCLLQBEKELVEPLTPSGQAPNQ 699
Qy 716 AQMRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPPKANKELD 775
Db 700 AFLRIKETEFKDRVLGSAFGTVYKGLWNPGENIRIPVAIKVLRATSPKVNQEVLD 759
Qy 776 EAYMAGVGSFYVSRILGICLTSTVOLVTLQMPYGCILLDHYVNRGLSGODLLNWCQOI 835
Db 760 EAYMASVDHPHVCRLGICLTSAVLVTLQMPYGCILLDHYVNRGLSGODLLNWCQOI 819
Qy 836 AKGMSYLEDVRLHRLDAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKWM 895
Db 820 AKGMYLEERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADKEYQADGGKVPKWM 879
Qy 896 ALESILRRRTHQSDVMSGYVYVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTI 955
Db 880 ALESILQWTVTHQSDVMSGYVYVWELMTFGSKPYDGIPIAREIPDLLEKGERLPPOPICTI 939
Qy 956 DVYIMVYKWMIDSECRPRELVSFBSRMARDQRFVYVQNEDLGPASPLDSTFYKSL 1015
Db 940 EYVIMILKWMIDSPSRPRELVSFBSRMARDQRFVYVQNEDLGPASPLDSTFYKSL 996
Qy 1016 EDDMGDLVDAEYLVQOQFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEE 1075
Db 997 SSD--DVDADEYLLPKRI-----NRQS----- 1020
Qy 1076 APRSLAPSEGAGSDVDFDGLMGAKGLOSLTHDPSPLORYSEDPTV-PLPSETGYV 1134
Db 1021 ---EPCIPPTGH-----PVRENSITLNRISDPTQNALEKDLGDH- 1056
Qy 1135 APLTCSQPEYVNPQDVRPQ-----PSPRE-----GPLP-AARFAGATLERAKTLSP 1181
Db 1057 -----EYVNPQGETSRSLSDIYNPNYEDLTGCGVPVSLSSQEAETNFSRPEYLNT 1107
Qy 1182 GKNGVWKDVFAFGAVENPEYLTQGGAAAPQHPPPAFSPAFDNLVYWDQDPPRGPAPPS 1241
Db 1108 NQNSL---PLVSSGSMDDPDY---QAG-----YQAFA-----LPQTGALTG 1142
Qy 1242 TFKGTPTAENPEYLG 1256
Db 1143 NGMFLPAENLEYLG 1157
```

RESULT 9

```
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
GN (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=90083234; PubMed=2687875;
Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
"Isolation and characterization of ERBB3, a third member of the
ERBB/epidermal growth factor receptor family: evidence for
overexpression in a subset of human mammary tumors.";
Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
[2]
SEQUENCE FROM N.A.
MEDLINE=90311312; PubMed=2164210;
Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
Todor G.J., Shoyab M.;
"Molecular cloning and expression of an additional epidermal growth
factor receptor-related gene.";
Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
[3]
SEQUENCE FROM N.A. (SHORT FORM).
TISSUE=Placenta;
MEDLINE=93282822; PubMed=7685162;
Kato M., Yazaki Y., Sugimura T., Terada M.;
"c-erbB3 gene encodes secreted as well as transmembrane receptor
tyrosine kinase.";
Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
-!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
SECRETED (SHORT FORM).
-!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
-!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
-!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
-!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M29366; AAA35790.1; -
EMBL; M34309; AAA35979.1; -
EMBL; S61953; AAB26935.1; -
PIR; A36223; A36223.
HSSP; P11362; 1FGK.
Genew; HGNC:3431; ERBB3.
MIM; 190151; -
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
Pfam; PF00069; pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SM00219; TyrKc; 1.
PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
```

KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 552 565
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT VARSPLIC 141 183
FT VARSPLIC 184 1342
FT CONFLICT 560 560
FT CONFLICT 1064 1064
FT SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
Query Match 34.3%; Score 2342.5; DB 1; Length 1342;
Best Local Similarity 39.8%; Pred. No. 1.le-119;
Matches 527; Conservative 188; Mismatches 459; Indels 151; Gaps 34;
QY 10 GLLALLPPGAA--STQVCTGTDMLKRLPASPEHLDMLRLHYQGVQVQGNLELYLPT 67
DB 11 GLLFSLARGSEVGNQAVCPGLTGLSVTGDENQVQTLKLYRCEVVGWGLEIVLTGH 70
QY 68 NASLSFLQDIQEVQGVVLIAHNOVQPLQRIVRGTQTFEDNYALAVLDNGDPLNNTT 127
DB 71 NADLSFLQIREVTGYLVVAMNEFTLPLPLNRVRVGTQVYDCKFAIFVM-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSLTEILKGGVLIQIORNPOLCYQDTILAKDIFHKNNQALALTIDTN 187
DB 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNKLCMDTIDWRDVRDRD----AEIVVKD 178
QY 188 RSRACHPCSPMCKGRGSESSDCQSLTRTVTCAGGC-ARCKGPLPFFNFTVSFMLRVPK 246
DB 179 NGRSCPPCHEVCKG-RCWPGSEDCQTLTKTICAPQCNGHCGPNP----- 223
QY 247 VSASHLE-----KHSCLACLHNSGICELHCPALVTYNTDTFESMPNPEGRTFG 298
DB 224 NQCHDECAGCGSGPDQDTCFACRHFNDSGACVPRCPQPLVYNKLTFOLEPNPHTKYQG 283

QY 299 ASCVTACPNYLYSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLRE 358
DB 284 GVCVASCPRNFV--VDQTSVCVRACPPDKRMEVD--KNGLAKHCEPCGGLCPKACEGTGSG--SR 339
QY 359 VRVTSANTQEFAGCKKI FGSFLAPLPESDPDASNTAPLQPEQLQVVFETLEEITGYLYI 418
DB 340 FQTVDSNIDGFCVCTKILGNLDFLITGLNGDPWHKI PALDPEKLVNFRVREITGYLNI 399
QY 419 SAMPDSLPLDSVFQNLQVIRGRILHNGAYS-LTQGLGISWLGRLSLRSLRGSGIALIHN 477
DB 400 QSWPPHMHNFSPVSNLTITIGRSLYNRGSLIMKLNLTSLGFRSLKEISAGRIYISAN 459
QY 478 THLCFVHTVPWDQLFRPHQALLHTA-NRPEDECVGEGELACHQLCARGHGWGPOPTQCVN 536
DB 460 RQLCYHHSNLNWTKVLGRPTTEERLDIKHNRRPRDCVAEGKVCPLCSSGGCGMGPQGCLS 519
QY 537 CSQFLRGQECVEECRVLQGLPREYVVARHCLCPHCEQPONGSVTCFPGPADQCVACAHY 596
DB 520 CRNYSRGGVCVTHCNFLNGEPREFEAHEACFSCHPECPMEGTATCNGSSGSDTCAQCAHF 579
QY 597 KDPFFCVARCPGKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEORA-- 654
DB 580 RQGPFCVSSCPHGVLG--AKGPIYKYPDVQNECRPCHECTCGCKGPELODCLQTLVLI 637
QY 655 --SPLTSIVSAVVGILLVVVLGVVFGILIKRROOKIR-KYMRRLLOETELVEPLTPSGA 711
DB 638 GKTHLTWALTVIAG--LVWIFMLGGTFLYWRGRIQNKRAMRYLERSGESIEPLDPS-E 694
QY 712 MPNQAQMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANK 771
DB 695 KANKVLARIFKETELRKVLGSGVFGTVHKGWVIEGESIKIPVCIKVIDESGRQSFQ 754
QY 772 EILDEAYVMAGVSPVSRLLGCLTSTVQLVTLQMLPYGCLLDHVRNRRGLASQDLLNW 831
DB 755 AVTDHMLAIGSLDHAHIVRLGLGCPGSSQLQVLYPLGSLLDHVRHQRGALCPQLLNM 814
QY 832 CHQIAGKMSYLEDRVLRVLRDLAARNVLKSPNVRKITDPCGLARLLDIDETEHADGKVP 891
DB 815 GVQIAGKMYLEEHGWHNRNLAARNVLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTP 874
QY 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTGAKPYDGI PAREITPDLLEKERLPQP 951
DB 875 IKWMALESIFPKYTHQSDVWSYGVTVWELMTGAEYPAGRLAEPDLDLEKERLAQOP 934
QY 952 ICTIDYIMVIMVCKMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPA---SPLDS 1008
DB 935 ICTIDYIMVIMVCKMIDENIRPTFKELANEFTMARDPPRYLYIKRES--GPGIAPGPEPH 993
QY 1009 TYRSLLLEDDMGDLVDAEYLVPOQGFPCPDPAAGAGGMVHHRRSSSTRSGGDLTLG 1068
DB 994 GLTNKKLEVELEPELDDLDLEAED-----NLATTLGSAULSLP 1034
QY 1069 LEP-SEEEAPRSLAPSEGAGSDVFDGDLGMNAKGLQSLPTHQ-PSPLORYSEDPTVPL 1126
DB 1035 VGTLNRPQSGQLLSPSGY-MPMNQNLGESCQESAVSGSSRERPRPSLH-----PM 1087
QY 1127 P-----SETDGYVA-----PLTCSQPEB-----VYNQPDVRPOPSPRE 1160
DB 1088 PRGCLASESGHVTGSEAELOEKYCMCRSRSRSPRGRGSAYHSQRHSLTPTVPLS 1147
QY 1161 GP-----LPAARPAGATLERAKTLPSP-GKNGVV-----KDVFAFGGAVENPEY 1202
DB 1148 PPGLEEDVNGVYMPDTHLKGTPSSREGTSSVGLSSVLGTGEDEED-----BEY 1199
QY 1203 LTPQGAAPQPPHPPAFSPAFDNLVYWD-----QDPPERGAAPPSTFKGPTAE 1250
DB 1200 MNRRRHSD-PPHPPRPSLLEELGYEYMDVGSLSASLGSTQSCPLHPVIMPAGTTPDE 1258
QY 1251 NPEYL 1255
DB 1259 DYEYM 1263

```
RESULT 10
ERB3 RAT
ID ERB3 RAT STANDARD; PRT; 1339 AA.
AC Q62759; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
DE ERB3.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=85212190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein.";
RL Gene 165:279-284(1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Koland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroligins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29339; AAC28498.2; -.
DR EMBL; U52530; AAC53050.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR PRINTS; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
```

```
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT TRANSMEM 644 662
FT DOMAIN 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
FT DISULFID 231 243
FT DISULFID 246 255
FT DISULFID 259 286
FT DISULFID 290 301
FT DISULFID 305 320
FT DISULFID 323 327
FT DISULFID 500 509
FT DISULFID 504 517
FT DISULFID 520 529
FT DISULFID 533 549
FT DISULFID 556 573
FT DISULFID 576 585
FT DISULFID 589 610
FT DISULFID 613 621
FT DISULFID 617 629
FT CARBOHYD 126 126
FT CARBOHYD 250 250
FT CARBOHYD 353 353
FT CARBOHYD 408 408
FT CARBOHYD 414 414
FT CARBOHYD 437 437
FT CARBOHYD 469 469
FT CARBOHYD 522 522
FT CARBOHYD 566 566
FT CARBOHYD 616 616
FT CONFLICT 1028 1028
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BPDF1E CRC64;

Query Match 33.3%; Score 2271.5; DB 1; Length 1339;
Best Local Similarity 40.0%; Pred No. 7.7e-116;
Matches 519; Conservative 168; Mismatches 433; Indels 177; Gaps 36;

Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLRLPASPETHIDMLRHLHYGQCQWQGN 59
Db 7 LQVLC---FLLSLARGSEMGNSQAVCPGTLNGLSVTGDADNQYQTLVKLYEKCEVVMGN 62

Qy 60 LELTYLPTNASLSFLQDIQIEVOGYVLIJAHNOVROVPLQRLIRIVRGTLQFDENYALAVLDN 119
Db 63 LEIVLTGHNADLSFLQIWIREVTVGYLVANNEFSVPLPLNLRVVRGTQYDVGKFAIFVN-- 120

Qy 120 GDPINNTTPTVTGASPGGLRELQLRLSLTEILKGGVLIQIORNPOLCYODTILWKDIFHKNQL 179
Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDMRDVRVVR 170

Qy 180 ALTLIDTNRSRACHPCSPMCKSGRCMGESSEDCSLTRTVCAGGC-ARCKGLPLPNNFTV 238
Db 171 GAELVVKNGANGCPPCHEVCCKG-RCWGPDPDCCQLTKTICAPQCNRCFCFNP----- 223

Qy 239 SFWLRLPKVKSASHLE-----KHSCLACLHPNHSGLCELHCPALVTYNTDFTESMPN 290
Db 224 -----NQCHDECAGGCGSQDTCFACRRFNDSGACVPRCPPEPLVYNKLTFOLEPN 275

Qy 291 PEGRYTFGASCVTACPVNYLSTVGSCTLVCPHLNQSVTAEDGTQRCCKSKPCARVCYG 350
Db 276 PHTKYQYGGVCVASCCHPNFV-VQDTFCVRACPPDKMEVD-KHGLKMCPECGGLCPKACEG 333

Qy 351 LGMEHLREVRAVTSANIQEFAGCKIFGSLAFLPSPDGPASNTAPLQPEQLQVFTLE 410
```


RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RL [7]
RP SEQUENCE OF 959-1078 FROM N.A.
RN STRAIN=Daekwanryeong;
RC MEDLINE=85137938; PubMed=2983232;
RX Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RA "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor,"
RL Nature 314:178-180(1985).
RL [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RN ANALYSIS.
RP MEDLINE=92038942; PubMed=1936959;
RX Raza E., Schejter E.D., Shilo B.Z.;
RA "Interallelic complementation among *DER/fib* alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases,"
RL Genetics 129:191-201(1991).
RL [9]
RP REVIEW
RP MEDLINE=97248481; PubMed=9094709;
RX Perrimon N., Perkins L.A.;
RA "There must be 50 ways to rule the signal: the case of the *Drosophila*
RT EGF receptor,"
RL Cell 89:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAP-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOTEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF052754; AAC08536.1; --
CC EMBL; AF052753; AAC08536.1; JOINED.
CC EMBL; AF052754; AAC08535.1; --
CC EMBL; AF052752; AAC08535.1; JOINED.
CC EMBL; K03054; AAA51462.1; --
CC EMBL; K03417; AAA51460.1; --

DR EMBL; K03416; AAA50965.1; --
DR EMBL; K03418; AAA51461.1; --
DR EMBL; AF109077; AAD26134.1; --
DR EMBL; AF109078; AAD26132.1; --
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; --
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; --
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; --
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; --
DR EMBL; AE003454; AAF46732.1; --
DR EMBL; K02293; CAA26157.1; --
DR EMBL; X78920; CAA55523.1; --
DR EMBL; X78918; CAA55521.1; --
DR EMBL; X78919; CAA55522.1; --
DR PIR; A00640; GQFFE.
DR HSP; P11362; IFGK.
DR FlyBase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_Ldomain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_Ldomain; 2.
DR PRINTS; PD00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30
FT CHAIN 31 1426
FT DOMAIN 31 868
FT TRANSMEM 869 889
FT DOMAIN 890 1426
FT DOMAIN 938 1198
FT NP_BIND 944 952
FT BINDING 971 971
FT ACT_SITE 1063 1063
FT MOD_RES 902 902
Query Match 27.6%; Score 1883; DB 1; Length 1426;
Best Local Similarity 32.1%; Pred. No. 9.8e-95;
Matches 461; Conservative 184; Mismatches 433; Indels 360; Gaps 43;
Qy 24 QVCTGDMKURLPASPEETHDMLRHLYQGCVVQGNLELYLPT-NASLSFLQDIQEVQG 82
Db 100 KICIGTKSRLSVPSNKEHHYRNLDRYTNCTYDGNLKLTLWLPNENLDSPLDNIREVTG 159
Qy 83 YVLIANNQVQVPLQRLIRIVRGTLF-----EDNYALAVLDNGDPLNTPPTVTCASPGGL 137
Db 160 YILISHVDVKKVFPKLIIRGTLFLSLSVEEEKYALFV-----TYSKM 203
Qy 138 RELQLASLTILKGGVLIQRNPOLCYQDITLWKIDIFHKNNQLALTLDITNRSRACHPCSP 197
Db 204 YTLIEPDLRDVLNGQVGFHNNYLCHWRTTQWSEIVSNGTDAYNYDFTAPERCPCKE 263
Qy 198 MCKSGRCWGSSEDCQSLTRTVCAAGCA--RCKGPLPFNNPTVSVFWLRVPKVSASHL--- 252
Db 264 SCTHG-CWGGPKNCKFKSLTCSPOCAGGRCYGPKP-----RECCHLFCA 308
Qy 253 -----EKHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTTGASCVTACP 307
Db 309 GGCTGPTQKDCIACKNFFDEAVSKBECPPMKRKNPTTYVLETNPEGKYAYGATCVKECP- 367
Qy 308 NYLSTDVGSCTLVCPLNHQBVTAEADGTQRCCKSCPCARVCYGLGMHLEHRAVTSANI 367

```
Db 368 GHLLRNGACVRSQPOKMDKGE-----CVPCNGCPKTCPCGVTVLH-----AGNI 414
Qy 368 QEFAGCKIFGSLAFIPESFDG--DPASNTA-----PLQEOLOVFETLEEITGYLYIS 419
Db 415 DSFRNCTVIGNIRILDQTFSGFDQVYANYTWGPRVPLDPEREVEFSTVKEITGYLNI 474
Qy 420 ANPDSLPLDSVFONLOVIRGRIHNGAY-SLTLOGLGISWGLRSURELGSGLALIHNT 478
Db 475 GTHPOPRNLSYFRNLETIRGOLMESFAALAIVKSSLSYLEMRNLKQISSGSSVVOHNR 534
Qy 479 HLCFVHTVPWDDLFRNPHQALLHTANRPEDECVGEGLACHQIOLCARGCHGPGPTQCVNCS 538
Db 535 DLQYVSNIRWPAIQKEPEQKVVWVWENLRADLCEKNTICSDQNEGCGWAGTDQCLTCK 594
Qy 539 QFLRQGECEVECRVLOGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKD 598
Db 595 NFNFGTCTIADCGYISNAYK--FDNRKTKICHECR-----TCNGAGADHCOECVHVD 646
Qy 599 PPFVCVARCP-----SGVK-----611
Db 647 GQHCVSECPKNKYNDRGVCRECHATCGCTGPKDTIGACTTCNLAIINNDATVKRCLL 706
Qy 612 -----PDLSPYMPIWKF--PDEGACOP-----CPI-----NCTH- 638
Db 707 KDDKCPD-CY--FWEYVHPQEGSLKPLAGRAVCRKCHPLCLBLCTNYGHEQVCKTHY 763
Qy 639 -----SC-----VDLDDKG-----647
Db 764 KRREQETECPADHYTDEBQECFQHPBCNGCTGPGADDCKSCRNFKLF DANETGPYV 823
Qy 648 -----CPAEQR-----ASPLTS-----IVSAVVGILLVV 671
Db 824 STMFNCTSKCPLMRHVNVQYTAIGPYCAASPPRSKITANLDVNMIFITGAVLVPTIC 883
Qy 672 VLGVVFGILIKRQOKIRKYT--MRLLOTELVELTPSGAMPQAOAMRILKTELKRV 729
Db 884 ILCVV--TYICQKQKAKKETVMTWALSGCDESEPLRFSNTGALCKURIKVDAELRG 941
Qy 730 KYLGSGAFVTYVYGIWIPDENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVS 789
Db 942 GVLGMGAFGRVYGVVPEGENVKIPVAIKELKSTGAESSEFLEAREAYIMASEEHVNL 1001
Qy 790 RLGLCLSTVOLVTOLMPYGCLLDVRNRENRGLSGQDLNMCQIAKMSYLEDRVLVH 849
Db 1002 KLLAVCMSSQMLITQLMPLGLGLDYVRNRDKIGSKALLNWSQTAKGMSYLEKRLVH 1061
Qy 850 RDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRBFTQS 909
Db 1062 RDLAARNVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGKMPDKWLALECIERNRVFTSKS 1121
Qy 910 DVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVTYIMVMVKWMIDS 969
Db 1122 DVMAFGVTIWELLTFGQRPHEINPAKIDPDLIEVGLKLEQPEICSLDIYCTLLSCWILDA 1181
Qy 970 ECRPFRELVSERFMRDQRFVVIQNEIDLQ--PASPLDSTFYRSLLDD--DMGDLV 1024
Db 1182 AMRPTFKLTVTFABFADPGRYLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPT 1234
Qy 1025 DAEYLVLPQOQFFCDDPAPGAGGVVHRRHSSTSRSGGDLTLGLEPSEERAP-----R 1078
Db 1235 DGSEAIAPDDYLOPKAAPGPS-----HRTDCT-----DEMPKLNRYCK 1273
Qy 1079 SPLAPSEGAGSDVFG-----DLGMAAKGLQLSLPHTHPSPLQRYSDPTVPLPSETDGYVA 1135
Db 1274 DPSNKNSTGDDERSSAREVGVGNLR-----LDLPVDEDDYLM 1312
Qy 1136 PLTCSPOEYVNPQDVRPQSPREGPLPAARPACATLERAKTSLSPGNKGVVQVFAFG 1195
Db 1313 P-TQCPGPNNNNNM-----NPNQNNMAVGAAGYM-----DLIGVPV 1350
Qy 1196 AVENPEYL-----TPOGGAPOPH-----PPPAFSP-AFDNLYYWD 1230
```

```
Db 1351 SVDNPEYLLNAQTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPSGTSSTSDHEYND 1408

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
ID ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodman R.G., Rottman F.M.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
processing and promoter insertion result in expression of an
amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSP; P11362; LFQK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transfaser; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.6%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 6.6e-88;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFDEEGACQPCPCINCTHSCVDLDDKGPAG 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCOLCHPNCTRGCKPGLEGCP--- 58
Qy 653 RASPLTSIVSAVV-GILLVWVGVFGILIKRQOKIRKYTRRLQLQETELVELPTPSGA 711
```



```
Db 59 NSGKTPSIAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRLLQERLEVEPLTPSGE 117
Qy 712 MPNQAMRILKETELARKVVLGSGAGFTYVKGHWIPDGNVKIPVAIKVLRNTPSKANK 771
Db 118 APNQALRLKETEFKKVVLGSGAGFTYVKGHWIPEGEKVPVPAIKELREATSPKANK 177
Qy 772 EILDEAYVAGVGVSVVRLGICLTSTVQLVTPGLDHPVRENRLGSGDGLLNM 831
Db 178 EILDEAYVAGVGVSVVRLGICLTSTVQLVTPGLDHPVRENRLGSGDGLLNM 237
Qy 832 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPARETIDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCWMIDSECRPRFRELVSFEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCWMIDSRPKFRELIAEFSKWARDPPRYLVIGDERHKLPSPTDSKF 417
Qy 1011 YRSLLDDDMGDLVDAEYLVPOQGFPCPDPAAGAGMWHRRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDEMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1071 PSEEEAPRPL-----APEGAGSDVFDGDLGMAKGLQSLTPHPSPLORYSEPTVP 1125
Db 450 -----SRTPLLSLSSATSNNSATCID-----RNGQGHVPREDVSFVORYSSDPTGN 495
Qy 1126 LPSET--DGVVAPLTCSPQEVYVQDVRPPOPPSPREGPLPAARPAATLERAKTLSPGK 1183
Db 496 FLEESIDDFL-----PAPEYVQ--LMPKES-----TAMVQ 526
Qy 1184 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPHPPAFSPAFONLY 1228
Db 527 NQIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTQSPKLVTFSSPY 578
Qy 1229 WQO-----DPPE-----RGAPPSTFKGTPTAENPEYLCGLDVP 1260
Db 579 WQSGNHQNLNDPQYQDFLFPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC F07535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erBB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain E54).
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erBB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84221957; PubMed=6328659;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erBA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
```

```
CC tyrosine phosphate.
CC DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02006; AAA42394.1; ALT_INIT.
CC EMBL; K01216; AAA42400.1; -.
CC PIR; A00644; TVYUH.
CC HSSP; P11362; LFQK.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
CC Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 2.1e-85;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 593 CAHYKDPFPCVACPSGVKPDLSYMPWPFDEEGACPCPINCTHSCVDLDDKGCPEAQ 652
Db 3 CAHFIDGPHCVKACPAVLGENDTL--VRKYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58

Qy 653 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQKIRKYTRRLQETELVEPLTPSGA 711
Db 59 NSGKTPSIAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRLLQERLEVEPLTPSGE 117
Qy 712 MPNQAMRILKETELARKVVLGSGAGFTYVKGHWIPDGNVKIPVAIKVLRNTPSKANK 771
Db 118 APNQALRLKETEFKKVVLGSGAGFTYVKGHWIPEGEKVPVPAIKELREATSPKANK 177
Qy 772 EILDEAYVAGVGVSVVRLGICLTSTVQLVTPGLDHPVRENRLGSGDGLLNM 831
Db 178 EILDEAYVAGVGVSVVRLGICLTSTVQLVTPGLDHPVRENRLGSGDGLLNM 237
Qy 832 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVP 297
Qy 892 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPARETIDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 952 ICTIDVYIMVKCWMIDSECRPRFRELVSFEFSRMARDPQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCWMIDSRPKFRELIAEFSKWARDPPRYLVIGDERHKLPSPTDSKF 417
Qy 1011 YRSLLDDDMGDLVDAEYLVPOQGFPCPDPAAGAGMWHRRHRSSTRSGGDLTLGLE 1070
```

```

Db 418 YETLMEEDMEDIVDAEVLVPHQGF-----NSPST----- 449
QY 1071 PSEEBAPRPL-----APSEGAGSDVFDGLGMAAGKQSLPTHDPSPLOQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHVPVREDSFVQRYSSDPTGN 495
QY 1126 LPSET--DGYVAPLTCSPQPEYVQPDVPPQPSREGPLPAARPAAGATLERAKTLSPK 1183
Db 496 FLEESIDQGL-----PAPEYVQ--LMPKKPSTAM----- 524
QY 1184 NGVVKDVFAP-----GGAVENPEYLTPOGGAAPQHPHPPAFSPAFD 1224
Db 525 --VQNIYNFISLTAISKLPMDSRVQNSHSTAVDNPYL-----NTNQSPKLAKTVEF 574
QY 1225 NLYVNDQDPPERGAPPSTFGTPTAENPEY 1254
Db 575 SSPYIQSGNHQ-----INLDNPDY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.";
RL Mol. Cell. Biol. 6:1751-1759 (1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA42401.1; -
CC PIR; A25231; TVFVEB.
CC HSP; P11362; IFGK.
CC InterPro: IPR000719; Euk pkinase.
CC InterPro: IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; TyrK; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 23.9%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 1.7e-81;

```

```

Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;
QY 593 CAHYDPPFCVACRPSGVKPDLSYMPIMKFPDQBEAGCQPCPINCTHSCVDLDDKGCPEAQ 652
Db 3 CAHFIDGPHCVKACPAVLGENDTL--VMKYADANAVCOLCHPNCNTRCCKCGPGLGECF--- 58
QY 653 RASPLTSIVSAVV--GILLVVVGVFGIILIKRQKIRKYTMRRLLQETELVELPLTPSGA 711
Db 59 NGSKTPTSAAGVGVGLLCLVVGIGLGLYLR--HIVRKRTLRLRLQERELVELPLTPSGE 117
QY 712 MNOQAMRLTKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANK 771
Db 118 APNQAHLRLTKETEFKVKVLGFGAGTGVYKGLWIPGEKVTIPVAIKEUREATSPKANK 177
QY 772 EILDEAYVNAVGSPPVSRLLGICLTSTVQLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV 831
Db 178 EILDEAYVNAVSDNPHVCRLLGLICLTSTVQLTQVLTQVLTQVLTQVLTQVLTQVLTQV 237
QY 832 CMQIAKGSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVP 891
Db 238 CVQIAKGMNLYLEERHMHVHRDLAARNVVKTPQHKITDFGLAKQLGADEKEYHAEGKVP 297
QY 892 IKWMALESILRRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQPP 357
QY 952 ICTIDVIMVWKCMWIDSECRPRFRELVESEFMRMARDPQRFVVIQ--NEDLGASPLDSTF 1010
Db 358 ICTIDVIMVWKCMWDADSRPKFRELIAEFSEKWARDPPRYLVIOQDERHKLSPDTSKF 417
QY 1011 YRSLLDEDDMDGLVDAEYLVPOQGFCCPDPAAGAGMVHRRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEVLVPHQGF-----NSPST----- 449
QY 1071 PSEEBAPRPL-----APSEGAGSDVFDGLGMAAGKQSLPTHDPSPLOQRYSEDPTVP 1125
Db 450 -----SRTPLLSLSLTSATSNNSATNCIDRNGG-----H----- 476
QY 1126 LPSETDGYVAPLTCSPQPEYVQPDVPPQPSREGPLPAARPAAGAT--LERAKTLSPGKN 1184
Db 477 -PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNIYVLSITASKLPIDSRVQN 527
QY 1185 GVVKDVFAPGGAVENPEYL 1203
Db 528 -----SHSTAVDNPYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax J., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha";
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND

```

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M20386; AAA48760.1; -;
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC SMART; SM00261; FU; 4.
CC
CC PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; PARTIAL.
CC PROSITE; PS00111; PROTEIN KINASE_DOM; PARTIAL.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Matches 304; Conservative 114; Mismatches 253; Indels 50; Gaps 14;
QY 8 RMGLLLALLPPGAA-----STQVCTGTDMKRLRASPETHLDMLRHLRYOGCOVVOGNLE 61
DB 13 RGAALVLLGLLGLVCAVEEKKVCOGTNNKLTQLGHVEDHFTSLQRMNCCVLSNLE 72
QY 62 LTYLPTNALSFLQDIQEQGVVLAHNOVROVPLQRLRIVRGTLQFEDNYALAVLDNGD 121
DB 73 IYVEHNRDLTLFKTIQEVAGIVLALANMVDVIPLENLQIIRGNVLYDSFALAVLSNYH 132
QY 122 PLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNQAL 181
DB 133 -MNTQ-----GLRELPMKRLSEILNGGVKISNNPKLNNMDTVLWMDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPCKGRCWSESSEDCSLTRTYCAGCA-RCKGPLFPNNFTVS 239
DB 183 TVLDFASNUSCPKCHPNCTEDHCWAGEQNCOTLTKVCAQOCGRCKGKVPSPD----- 237
QY 240 FWLRVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTVNTDTFESMNP 291
DB 238 -----CHNQCAAGCTGPRSDCLACRKRDDATCKDTCPLVLYNPTTYQMDVNP 288
QY 292 EGRYTFGASCVTACPNYLSLTVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARCYGL 351
DB 289 EGKYSFGATCVRECPHNYVVDHGSVCVRSCTDITYEV-BENGVRKCKCDGLCSKVCGI 347
QY 352 GMEHLREVRAVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEOLOVFELEE 411
DB 348 GIGELKGLISINATNIDSFNCTKINGDVSIIPLVAFLDGDAFTKTLPLDPKKLDVFTVKE 407
QY 412 ITGYLYISAWPDSLPLSVFQNLQVIRGIRILHNGAYSLTLQGLGISWLGSLRLSRELGSGL 471
DB 408 ISGFLLIQAWPDNATDLYAFENLEIRGTQKHQGYSLAVVNLKIOSLGLRSLEISDGD 467
QY 472 ALIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHGWGP 531
DB 468 IAIMKNKLCYADTMNRSLFATQSOKTKIIQNRKNKNDCTADRHVCDPLCSDVGCWGP 527
QY 532 TQVNCVQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPONG---SVTCFGPEAD 588
DB 528 FHCFSRFFSROKECVKQCNIIQGEPEFERDSKLPCHSECLVQNSTAYNTTCGSGP 587
QY 589 QCVACAHYKDPFPFCVARCPSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGC 648
DB 588 HCMKCAHFIDGPHCVKACFAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGPGLEG 646
QY 649 PAEQRASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLT 707
DB 647 P---NGSKTPSIAAGVVGGLCLLVVVGILGILYLRRL-HIVKRTLRRLIQERELVEPLT 702
QY 708 P 708
DB 703 P 703

Search completed: July 22, 2003, 08:45:17
Job time : 21.2793 secs

Query Match 22.1%; Score 1510; DB 1; Length 703;
Best Local Similarity 42.2%; Pred. No. 7.3e-75;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-210-224-14
Perfect score: 6827
Sequence: 1 MELAAALRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6173	90.4	1259	6	O18735
2	3057	44.8	1209	11	Q9GX70
3	3028	44.4	1210	11	Q9EP98
4	2652	38.8	1165	13	Q9VH40
5	2631.5	38.5	1137	13	Q9W6F6
6	2238	32.8	1328	13	Q9W754
7	1953.5	28.6	1433	5	Q9BIH9
8	1752.5	25.7	419	4	Q9UK79
9	1739	25.5	367	11	Q8R2X1
10	1720	25.2	729	15	Q8E712
11	1718	25.2	567	15	Q8E714
12	1697.5	24.9	412	4	Q8WYV0
13	1655.5	24.2	962	15	Q64895
14	1645	24.1	545	15	Q85468
15	1419.5	20.8	655	11	Q9WVFS
16	1403.5	20.6	643	11	Q9ERV6

17	1211	17.7	1193	5	Q9Y1X8
18	1157.5	17.0	1368	5	Q23821
19	1142	16.7	1717	5	Q26566
20	1041	15.2	527	13	Q90836
21	973.5	14.3	599	13	Q9PSH2
22	914.5	13.4	478	11	Q9ESE0
23	906	13.3	165	4	Q14256
24	887	13.0	176	11	Q923V5
25	806.5	11.8	346	13	P11776
26	778	11.4	435	5	Q8SZW1
27	754.5	11.1	311	13	Q9P162
28	707.5	10.4	1362	13	Q9PVZ4
29	678	9.9	1671	5	Q9NUV5
30	658.5	9.6	1368	13	Q8UW85
31	638	9.3	1418	13	Q93457
32	636	9.3	331	4	Q9BUD7
33	625	9.2	1358	13	Q73798
34	623.5	9.1	1472	5	Q9U5A8
35	620.5	9.1	1369	13	Q8UW86
36	604	8.8	1412	13	Q8UW84
37	601	8.8	149	6	Q9BG66
38	596	8.7	987	11	Q9LYM0
39	596	8.7	1371	11	Q9QVW4
40	593	8.7	987	11	Q99MR2
41	588.5	8.6	1418	13	Q8UW83
42	587.5	8.6	1036	4	Q07912
43	587.5	8.6	1055	11	Q54967
44	586.5	8.6	1091	4	Q9DMQ4
45	584.5	8.6	2144	5	Q9VD94

ALIGNMENTS

RESULT 1

O18735	PRELIMINARY;	PRT;	1259 AA.
ID	O18735;		
AC	O18735;		
DT	01-JAN-1998 (Tremblrel. 05, Created)		
DT	01-JAN-1998 (Tremblrel. 05, Last sequence update)		
DE	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Brb-2.		
(Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]_TaxID=9615;		
RP	SEQUENCE FROM N.A.		
RA	Yokota H.;		
RT	"cDNA cloning of erbB-2 from canine mammary gland."		
RL	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB008451; BAA23127.1;		
DR	HSSP; P11362; IFGK.		
DR	InterPro; IPR002048; EF-hand.		
DR	InterPro; IPR000494; EGFR_L_domain.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	InterPro; IPR004019; YLP_motif.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00069; pkinase; 1.		
DR	Pfam; PF01030; Recep_L_domain; 2.		
DR	Pfam; PF02757; YLP; 2.		
DR	ProDom; PD000001; Euk_pkinase; 1.		
DR	SMART; SM00261; FU; 3.		
DR	SMART; SM00219; TyrKc; 1.		
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
KW	ATP-binding; Transferase; Tyrosine-protein kinase.		
SEQUENCE	1259 AA; 137989 MW; E37364D49C4CD46 CRC64;		

Query Match		90.4%	Score 6173;	DB 6;	Length 1259;
Best Local Similarity		90.0%	Pred. No. 0;		
Matches 1146;		Conservative 40;	Mismatches 60;	Indels 28;	Gaps 4;
QY	1	MELAAACRGLLLALLPPGAASTQVCTGDKMLRASPETHLDMLRHLYQGQVQGNL	60		
DB	1	MELAAACRGLLLALLPSGAAGTQVCTGDKMLRASPETHLDMLRHLYQGQVQGNL	60		
QY	61	ELTYLPTNASLSFLQDIQEVQGVVLAHQVROVPLQRLIRVGTOLFDNYALAVLDNG	120		
DB	61	ELTYLPANASLSFLQDIQEVQGVVLAHQVROVPLQRLIRVGTOLFDNYALAVLDNG	120		
QY	121	DPUNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFKHNQLA	180		
DB	121	DPLEGGIPAPGAAGGLRELQRLSLTEILKGGVLIQRNPOLCHQDITLWKDFKHNQLA	180		
QY	181	LTLIDNTRSRACHPCSPMKSGSCWGESSEDCSLTRTVCCAGCARCKGPLPNNVTVSF	240		
DB	181	LTLIDNTRSFACPCSPACKDAHCWASSGDCOSLTRTVCCAGCARCKGPQ-----	232		
QY	241	WLRVPKVSASHLE-----KHSDCIACILFHNHSGICELHCPALVTYNTDTESPNPPE	292		
DB	233	-----TDCHECOAGCTGPKHSDCLACILFHNHSGICELHCPALVTYNTDTESPNPPE	286		
QY	293	GRYTFGASCVTACPNYVLTVDGSCITLVCPLHNOEVTAEQTCRCEKSKPCARVCYGLG	352		
DB	287	GRYTFGASCVTSCPNYVLTVDGSCITLVCPLNNOEVTAEQTCRCEKSKPCARVCYGLG	346		
QY	353	MEHLREVRVTSANIOEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEI	412		
DB	347	MEHLREVRVTSANIOEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLRVFEALEEI	406		
QY	413	TGYLYTSAMPDLSLDFSVQNLQVIRGRILHNGAYSILTQGLGISWGLSLRSLRELGSGLA	472		
DB	407	TGYLYTSAMPDLSLPLNSVQNLRVIRGRVLHDGAYSILTQGLGISWGLSLRSLRELGSGLA	466		
QY	473	LIHNTHLCFVHTVPMDQLFRNPQHALLTANRPEDECVEGGLACHOLCARGHCWPGPT	532		
DB	467	LIHNRNARLCFVHTVPMDQLFRNPQHALLHSANRPEBECVGEGLACYP-CAHGHGWCWPGPT	525		
QY	533	QCNCVSQFLRGQCEVCECRVLQGLPREYVNAHCLCHPECOPONGSVTCFGEADQOVA	592		
DB	526	QCNCVSQFLRGQCEVCECRVLQGLPREYVNDRYCLPCHSECQPNQNGSVTCFGEADQOVA	585		
QY	593	CAHYKDPPFCVACPSGVKPDLSYMPIWKFPFBEGACQPCINCTHSCVDLDDKGCPAEQ	652		
DB	586	CAHYKDPPFCVACPSGVKPDLSFMPWKFADESGTCQPCINCTHSCADLDEKGCPAEQ	645		
QY	653	RASPLTSIVSAAVVGILLVVLGVVFGILIKRROQKIRKYTMRLLOETELVEPLTPSGAM	712		
DB	646	RASPVTSIIAAVVGILLVAVVGLVGLILIKRRRQKIRKYTMRLLOETELVEPLTPSGAM	705		
QY	713	PHQAQWRIILKETLRKVKVGLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE	772		
DB	706	PHQAQWRIILKETLRKVKVGLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE	765		
QY	773	ILDEAYMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENHRLGSGQDLLNMC	832		
DB	766	ILDEAYMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCILLDHVREHRLGSGQDLLNMC	825		
QY	833	MQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVP	892		
DB	826	VQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVP	885		
QY	893	KWMALESILRRRTHOSDWSVGVVYVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPI	952		
DB	886	KWMALESIPRRRTHOSDWSVGVVYVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPI	945		
QY	953	CTIDVYIMVKWCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYR	1012		
DB	946	CTIDVYIMVKWCMIDSECRPRFRELVAEFSRMARDPQRFVVIQNEEDLGPASPLDSTFYR	1005		
QY	1013	SILLEDDMDGLVDAEEYLVPOQFFCDDPAPGAGMWHRRHSSSTRSGGDLTLGLEPS	1072		

Query Match 44.8%; Score 3057; DB 11; Length 1209;
Best Local Similarity 48.8%; Pred. No. 2.7e-22;


```
DR PROSITE; PS00130; CYTOCHROME C; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134940 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.4%; Score 3028; DB 11; Length 1210;
Best Local Similarity 48.3%; Pred. No. 4.2e-220;
Matches 621; Conservative 170; Mismatches 363; Indels 132; Gaps 25;

QY 11 LLALPPGAA--STVCTGTDMKRLRLPASPETHLMDRLHLYGCGQVQGNLELTPLTN 68
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 ASLSFLQDQEVGVYLAHNOVROVPLORLVRGTQLPEDNYALAVLQNGDPLNNTTP 128
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDI----FKNNQLALTLI 184
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 -YGTNRITGLRELPMRLQELICAVRFSNNPILCNMDTIQWRDIQNVFMSNMDL--- 180
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 DTNRSRACHPCSPMKGSRGWSGESSDDCQLTRTVGAGCA-RCKGPLPNNFTVSFWLR 243
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 -QSHPSKCPKDCPCNGSCWCGGEBNCQKTKIICAQCSHRCGRSPSD----- 230
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 244 VPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPGRY 295
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 -----CCHNQCAAGCTGPRSDCLVCQKFODEATCKDTCPLMLNPTTYQMDVNPGEKY 285
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 296 TFGASCVTACPNYLSLTDVGSCTLVCPLNHQEYTAEDGTQRCCKSKPCARVCYGLGMEH 355
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 SFQATCVKCPRYVYVDHGSCVRCGPDYEV-EEGIRKCKKCGPCRVKCVNGIGIGE 344
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 356 LREVRVAVTSANIOEPAGCKIFGSLAPLPSFGDPSASNTAPLQPSOLQVFLEITGY 415
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 FKDTLSINATNTHKFKYCTAISGLDLHLPVAFKGDSPTRTPPLDPRELEILTKVKEITGF 404
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 LYSISAMPDSLPLDSVFQNLQVIRGLIHNGAYSLTGLGIGSLGRLSRELGSGLALIH 475
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 405 LLIQAMPDNTDUAHENLEIIRGTQKHQGFSLAVVGLNITSLGRLSLEISDGDVVIS 464
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 HNTLFCVHTVPMDQLFRNPQALHTANRPEDECYGEGLACHQLCARGHCGPGTQCV 535
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 465 GNRNLVANTINMKLLFGTBNQTKIMNNAEKDCAVNHVNCPLSCSEGCWGPEDCV 524
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 536 NCSQFLRGQCEVCEVQLQGLPREYVNAHCLCHPECPQNGSVTCFGEADQCACAH 595
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 525 SCQVSRGRCVCEKCNILEGEPREFVENSECICHPCECLPQAMNITCTGRPDNCIQCAH 584
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 596 YKDPFPCVARCPSGVKPDLSYMPIKFPDBEGACQPCPINCTHSCVDLDDKGCAPQORAS 655
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 YIDGPHCVKTCFAGINGENNTL-VKVDANNVCHLCHANCTYGCAGPGLQGEVWPSPG 643
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 656 PLTISVAVVGIILLVVLGVVFGI-LIKRQOKIRKYTMRLRLOETELVEPLTPSGAMPN 714
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 644 KIPSIATGIVGLLFIW-VALGIGLFMRRRHVRKTRLLRLOERELVEPLTPSGEAPN 702
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 715 QAQMRLKETELRKVKVLGSGAGTYVKGWIIPDGENVKIPVAIKVLRENTSPKAKEIL 774
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 703 QAHLRLTKETEFKXIKVLGSGAGTYVKGWIIPDGENVKIPVAIKELREATSPKAKEIL 762
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 775 DEAYVMAGVSPVSVRLGICLTSTVOLVTQLMPYGCILLDHVRNRLGSGQDLLNCWQ 834
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 763 DEAYVMASVDNPNVCHRLGICLTSTVOLITQLMPYGCILLDYVRHNDKNGSQVLLNCWQ 822
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 835 IAKGMSYLEDLVRLVHRDLAARNVLVKS PNHVKITDFGLARLIDIDETEHADGCKGKPIKW 894
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 823 IAKGMNLEEDRLVHRDLAARNVLVKT PQHVKITDFGLAKLGAEEKEYHAEGKVPKW 882
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 895 MALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQDPIC 954
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
```


[illegible]

Qy	952	ICTIDVYIMVKWCMIDSECRPFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFY	1011
Db	934	ICTIEVYMIILKCMIDPSSRPFRFRELVSFEFSQWARDPSYLVIQG---NLPSPSDRRLF	990
Qy	1012	RSLLDDMDGDLVDAEYLYVPOQGFCCPDPAAGAGMGVHHRSSSTRSGGDLTLGLEP	1071
Db	991	SRLLSSDD--DVVDADAEYLL-----RYKRIN-RQGS-----	1018
Qy	1072	SEEEAPRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOBYSEDPV-PLPSET	1130
Db	1019	-----EPCIPNGH-----PVRENSIALRYISDPTQNALEKDL	1051
Qy	1131	DGVVAPLTCSPQPEYVYVQDVRPQP-----PSPRE-----GPLP-AARPACATLERAK	1177
Db	1052	DGH-----EYVYVQSPSETSSRLSDIYNPNYEDLTDGCVLSLSSGEAETNSRPE	1101
Qy	1178	TLSPGKNGVVKVFAFGGAVENPEYITPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERG	1237
Db	1102	YLTNTQNSL---PLVSSGSDDDPDY---QAG-----YQAAF-----LPQTG	1136
Qy	1238	APPSTFGKTPTAENPEYVGL	1257
Db	1137	ALTNGMGFLPAEAENLEYVGL	1156

RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Receptor tyrosine kinase (Fragment).

GN ERBB4.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

XP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HINDRAIN;

RX MEDLINE=99263203; PubMed=10328884;

RA Dixon M., Lumsden A.;

RT "Distribution of neuregulin-1 (nrgl) and erbB4 transcripts in embryonic chick hindbrain.;"

RL Mol. Cell. Neurosci. 13:237-258 (1999).

RL EMBL; AF121963; AAD31764.1; --

DR HSSP; P11362; 1FGK

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR000719; Euk_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001368; TNFR c6.

DR InterPro; IPR001245; Tyr_pkinase.

DR InterPro; IPR004019; YLP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF000659; pkinase; 1.

DR Pfam; PF01030; Recep_L_domain; 1.

DR Pfam; PF02757; YLP; 2.

DR PRINTS; PR00109; TYRKINASE.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TyrKC; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.

DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.

DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.

KW Kinase; Tyrosine-protein kinase.

FT NON_TER 1

SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 38.5%; Score 2631.5; DB 13. Length 1137.

Query Match 38.5%; Score 2631.5; DB 13; Length 1137;

Best Local Similarity 46.2%; Pred. No. 46-190; Matches 533; Conservative 166; Mismatches 355; Indels 99; Gaps 27;	
QY 161 LCYQDTILWKDIFHKNNQALALTLDTNRSRACHPCSPMKGSRGWESSEDQSLTRTV 220	DB 1199 NPEYLTPOGGAPOPHPPAPSPADNLYYMDODPERGA--PPSTFKGTP----- 1248
DB 3 LCFADTHIQDVRPNWASNFLLVPTNGSSGCRCHKCTG-RCWGPTENHCQTLLTKVC 61	DB 1050 NAEYL-----KNLPEKAKKAPDNDPMNHSUPPRSTLQHPDYLQVEYSTKYFKQNG 1101
QY 221 AGGC-ARCKGRLPFFNNFTVFWLRVPKVSASHLE-----KHSDCLACLHFNHSGICE 271	QY 1249 -----AENPEYL 1255
DB 62 AEQDCRCVGPY-----VS-----DCCHRECAAGCGSPKDDCFACMFNDSGACV 107	DB 1102 RIRPIVAENPEYL 1114
QY 272 LHCALVYNTDTFSPMPNPEGRYTFGASCVTACPNYLSLTDVSGCTLLVCPLNHQEVTA 331	RESULT 6
DB 108 TOCPOTFVYNPTTFOLEHNNHAKYTYGAFCKKCPHNFV-VDSSSCVACRACSPSKMEV-EE 165	ID P79754 PRELIMINARY; PRT; 1328 AA.
QY 332 DGTQCEKSRPCARVCYGLGMEHLREVRVANTSANIOBAGCKKIFGSLAFLPDFDGP 391	AC P79754;
DB 166 NGIKWCKPCTICPKACDGIPTGSLVSAQTVDSSNIDKPICTKINGNLIFLVTGIHGD 225	DT 01-MAY-1997 (TRENBLrel. 03, Created)
QY 392 ASNTAPLOPEQLQVETLEETIYLYISAWPDSLPLDSVFNQLQVIRGILHNGAYSJTL 451	DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DB 226 YHTIAAINPEKLNIFQTVREITGYLNIQSWPENMTDFVSNLVTIGGRALYSGLSLLIL 285	DE Erbb3.
QY 452 QGLGSIWGLSLRLRELGGSLAIHHNTHLCFVHTVPWDLFRNPHQALLHTANRDEBCV 511	OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
DB 286 KOOGITSLQFQSLKQISAGNIYITDINSNLCTYHTVNTSLFSTPSQKTVIHRNKAENCT 345	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
QY 512 GEGLACHOLCARGHCWGPGPTOCVNCOSFLRQCEVEECVLOGLPREYVNAHCLPCHP 571	OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
DB 346 ADGMVNCLECSDDGCGPGDCLSKCRPIRORTCIESCNLVDFGFRFANGSVCMCEDP 405	OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
QY 572 ECOP-QNGSCTFGPEADQVACAHYKDPFFCVARCPGKVPDLSPMPTWKPDPDEGACQ 630	OC Tetraodontidae; Takifugu.
DB 406 QCEKEMDNWITCYGPGDHCTCFHFKDGPNCVCKCPDLOGANSF--IFKYADEDRCH 463	OX NCBI_TaxID=31033;
QY 631 PCPINCTSHCVDLDDKGC-----PAQRASPLTSIVSAVV-GILLVVVLGVVFG 678	RN [1]
DB 464 PCHNPCTOCGRFASHDCIYYPWTQRSTLPOHAR-TPL--IAAGVIGGLFIIVIMGLTFA 520	RP SEQUENCE FROM N.A.
QY 679 ILIKRRQOKIRKYTWRRLLQETELVEPLTPSCAMPNOQMRLKETELRKVKVLGSGAGF 738	RA MEDLINE=9917347; PubMed=10077531;
DB 521 VTVRRKSK-KKRAURREL-ETELVEPLTPSGTAPNOAQLRLKETELRKVKVLGSGAGF 578	RA Gelliner K., Brenner S.;
QY 739 TVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPVSVRLGLICLTS 798	RT rubripes.";
DB 579 TVYKGIWPEGETVKIPVAIKILNETTGPKANVEFMDREALINASHDHPHLVLLGVCLSP 638	RL Genome Res. 9:251-258 (1999).
QY 799 TVQVLTQMPYGCCLLDHVRNRLGSGDILLNWCQIAKMSYLEDRVLVHRDLAARNVL 858	DR EMBL; AF056116; AAC34391.1; --
DB 639 TIQLVTQMPHGCCLLDYVHEHKDNIQSOLLNWCQIAKMMYLEERRLVHRDLAARNVL 698	DR HSSP; P11362; IFKGK.
QY 859 VKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALESLRFRFTQSDVMSVGVTV 918	DR InterPro; IPR000494; EGFR_L_domain.
DB 699 VKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALECHYRKRFTQSDVMSVGVTV 758	DR InterPro; IPR000719; Euk pkinase.
QY 919 WELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDYIMVMKCMWIDSCRPRFREL 978	DR InterPro; IPR001245; Tyr pkinase.
DB 759 WELMTFGKPYDGIPTREIPDLLEKGERLPQPPICTIDYIMVMKCMWIDASRPFKREL 818	DR Pfam; PF00757; Furin-like; 1.
QY 979 VSEFSRMARDPQRFVVIQNEH-LGPASPLDSTFYRSLLEDMDGDLVAEEVLPVQQGFF 1037	DR Pfam; PF01030; Recep_L_domain; 2.
DB 819 AAEFSRMARDPQRYLYIQGDDRMKLPSPNDSKFFQNLDEEDLEDMDMAEEYLPV-QAFN 877	DR ProDom; PD000001; Euk_pkinase; 1.
QY 1038 CPDPAPGAGGVHRRHSSTSGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDG 1094	DR SMART; SM00261; FU; 3.
DB 878 IPPPIYTRTRIDSNRNFVYRDGGYAAEQGV-PMPYRAPGCIIPAPVAQATAEIFED 936	DR SMART; SM00219; TyrKc; 1.
QY 1095 DLGMAAGKLOSLPHTDPSLQRYGEDPTVPLPS-----ETDGVAPLTCSPQPEYVN 1147	DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DB 937 TCCNGTLRKQVATLAKESDSTQRYSDPTVFIPERVIRGELEDGDMTPMRDKPTDYLN 996	DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
QY 1148 QPDVVRPQPPSPREGPLPAA-RPAGATLERAKTLSPGKNGVKDVF-----AFGAVE 1198	KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;	
Query Match 32.8%; Score 2238; DB 13; Length 1328;	
Best Local Similarity 39.3%; Pred No. 3.1e-160;	
Matches 515; Conservative 153; Mismatches 411; Indels 230; Gaps 32;	
QY 9 WGLLLALLPP--GAASTQ-----VCTGDMKRLPASPETHLDMLRHLVYOGCVVQGNLEL 62	QY 9 WGLLLALLPP--GAASTQ-----VCTGDMKRLPASPETHLDMLRHLVYOGCVVQGNLEL 62
DB 4 WRLILMCVASLRRAASSQTQEAQVCPQTQGLSSGTSGQENQYNLNRDVKYKGCBIINGNLEI 63	DB 4 WRLILMCVASLRRAASSQTQEAQVCPQTQGLSSGTSGQENQYNLNRDVKYKGCBIINGNLEI 63
QY 63 TYLPNASSLFLDIOEQVGVYLIHNNQVRQVLPQRLRIVRGTQLPEDNYALAVLDNGDP 122	QY 63 TYLPNASSLFLDIOEQVGVYLIHNNQVRQVLPQRLRIVRGTQLPEDNYALAVLDNGDP 122
DB 64 TQIESNWDPSFLKTIETRETVGYLIAMNHFQEIPLQGLRVIRGNSLYERRFALSVELN--- 120	DB 64 TQIESNWDPSFLKTIETRETVGYLIAMNHFQEIPLQGLRVIRGNSLYERRFALSVELN--- 120
QY 123 LNNTPTVTVGASPGGLRELQRLSLTEILAGGVLIQONPOLCYODTILWKDIFHKNNQAL 182	QY 123 LNNTPTVTVGASPGGLRELQRLSLTEILAGGVLIQONPOLCYODTILWKDIFHKNNQAL 182
DB 121 ----YPKDG--PSGLNQGLMNLTEILDGGVQIINNKYLRYGFWYWRDII--RNNDAPIE 173	DB 121 ----YPKDG--PSGLNQGLMNLTEILDGGVQIINNKYLRYGFWYWRDII--RNNDAPIE 173
QY 183 LIDTNRSRACHPCSPMKGSRGWESSEDQSLTRTVCAAGC----- 224	QY 183 LIDTNRSRACHPCSPMKGSRGWESSEDQSLTRTVCAAGC----- 224
DB 174 IQFNGERGVCVH---KSC-GNYCWGPGKQCCQLITKTVCAQCNDRCFCGTSRDCCHIECA 229	DB 174 IQFNGERGVCVH---KSC-GNYCWGPGKQCCQLITKTVCAQCNDRCFCGTSRDCCHIECA 229
QY 225 ARCKGRLPFFNNFTVFWLRVPKVSASHLEKHSKSDCLACLHFNHSGICELHCPALVYNTDT 284	QY 225 ARCKGRLPFFNNFTVFWLRVPKVSASHLEKHSKSDCLACLHFNHSGICELHCPALVYNTDT 284
DB 230 AGCKGPL-----DTDCFACRLFNDSGACVPCQPTLIYNKQT 266	DB 230 AGCKGPL-----DTDCFACRLFNDSGACVPCQPTLIYNKQT 266

```
QY 285 FESMPNPGRYTFGASCTVACPYNYLSTDVGCTLVCPHLNOEVTAEQDQOR-CEKCKSP 343
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 FQMETNPNKAYQYGVCSQCFTHFV-VDGSSCVSVCPDPKMEV--ERGSRQCELCSEL 323
QY 344 CARVCYGLGMEHLREVRVAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOL 403
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 324 CPKCEGTGAEE--QRQTVDSNIDISFNCTKIQGLSLHFLVTGILGDDPKNVPDADAKL 380
QY 404 QVFETLEEITGYLYISAMPDLSVFQNLQVIRGLHNGAYSILTLQGLGISWGLGRS 463
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 381 EVFRTVREITDILNQSPKELNDLSVFSSLTIIQGRSLFKRFSLMVMREIPTLTSGLRS 440
QY 464 LRELCSGLALHHTHLCFVHTVPWDQLPFRNH-QALLHTANRPEDECVGEGLAGHOLCA 522
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 LREISDGSVYISQNAHLCHYHTVNTQLFRGRVRANSLSNRPMACVADVADGRVCDPLCS 500
QY 523 RGHCHGPGTQCVCNSQFLRGQECVEECRVLOGLPREYVNAH-CLPCHPECPQNGSVT 581
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 501 DSGCMGPGDQCLSCRNYSRHGTCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGKAS 560
QY 582 CFGPEADOCVACAHYKDPFVCVARGSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCV 641
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 561 CTGPGADECMACTKRDGPYCNSSCPAGVN-DEKGLIFKFPNREGHCPECHQNCCTQGS 619
QY 642 DLDKGCPEAQRASPLTSIVSAVGLLVVVLGVVP-----GILIKRRQOKIRKYTMR 694
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 GPLNDCC--LEAARLTISSGQITGALGVPAGLIFCLVLFGLMLYHRLAIRRRKRAM 676
QY 695 RLLOETELVEPLTPSGAMPNOAQRILKETEELKVKVLGSGAGFTYKGIWIPDGENVKI 754
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 677 RYLESGESFEPUGP-GEKGTKVHARILKPSDLURKIKPLGSGVFTGSKGFVIPEGETVKI 735
QY 755 PVAIKVLRENTSFKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVQLVTQMPYGCCLID 814
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 736 PVAIKTIQDSSGRQFTTEITDHLSSMGSLDHPYIVRLGICPGTCLQVLTQLSSHGSLLE 795
QY 815 HVRENRLGSLQDLNWCQIAKHSYLEDLVLRDLAARNVLKPSNHNKXITDFGLAR 874
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 796 HIRQHTSLDQRLNWCQIAKGMYYLEEHRVHKNLAARNILLKNDYQVQISDYGVAD 855
QY 875 LLDIDETEYHADGGKVPKIMMALESILRRRFTHQSDVWSYGVTVVWELMTFFGAKPYDGIPA 934
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 856 LLYPDKKYVYSETKTPKIMMALESILFRYTHQSDVWSYGVTVVWMSFSGAEPYASQP 915
QY 935 REIPDLLEKGERLPPICTIDVYIMVKCMIDSECRPRFRRELVSFSEFMRMARDPQRFV 994
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 916 QEVPSVLEKGERLSQPAICTIDVYVMVKCMIDENIRPTFKELASDFRWARDPPIYLV 975
QY 995 IQNEDLGPASPLDSTFYRSLLEDDEDDMDGLVDABEYLVPOOGFFCPDPAPGAGMVHHRH 1054
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 976 IRMEG-----EDSGMGEPL-----R 990
QY 1055 SSSTRSGGDLTLGLEPSEERAPRPLAPSEGAGSDVFDGLGGMG---NAKGLQSLPHD 1111
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 991 RGSER---GLLEADLEDEEE-----GLGDRFATPQLQSPSPWS 1026
QY 1112 PSPLO-----RYSEDTPVLPSETDGYVAPLTCSPQ- EYVNO----- 1148
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1027 TSPSQINVMWMTQLRYD-----FAVSQGHIGYLPMSPSVDITRQLWYORSRLSSVR 1080
QY 1149 --PDVRPQPPSPREGPL--PAARPAQATLERAKTLSPGKNGVVKVDFAPGGAVENPEYLT 1204
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1081 TLPDRSAFRSREAELEDCEGACQACAGIFRVR-----FGSERGN----- 1118
QY 1205 POGGAAPQHPPPAPSPAFDNLYYWDQPPPERGAPPSFTKGTPTAENPE 1253
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1119 FOGG-----QQRKLSTASSPSPSKTWADEEDE 1146
```

RESULT 7

Q9BIH9

PRELIMINARY; PRT: 1433 AA.

AC Q9BIH9;

```
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR
CS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RP SEQUENCE FROM N.A.
RC STRAIN=SUJ;
RA Lyckett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; IFGK.
Ls InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D8896724F07 CRC64;

Query Match 28.6%; Score 1953.5; DB 5; Length 1433;
Best Local Similarity 31.8%; Pred. No. 1.2e-138;
Matches 465; Conservative 195; Mismatches 394; Indels 407; Gaps 38;

QY 26 CTGDMKRLRPASPTHLDMRLHVGCGVQVGNLELTYPNINASLFLQDIOEVGYVL 85
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 CIGTNGRMSVPANREYHYKNLRDRYTNCTYVDGNLEITWITQNTIDLNFLQHIREVGYVL 60
QY 86 IAHNOVROVPLQRLRVGTQLF-----EDNYALAVLDNGDPLNLTTPVTGASPGGLREL 140
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 ISLYDLQVILPRLQIIRGRTTFKLNKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSITEILKGGVLIQRNPQLCYQDTILWKDI-PHKNQALALTLIDTNKRACHPCSPMC 199
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 ELPALRDILGSGVGFNNYLNCHKMSINWEEILLAPQTSMQYTFNFSSEPERVCPCHPSC 164
QY 200 KGSRCWGSSESDCQSLTRTVCAAGCA--RCKGPLPNNFTVSPWLRVPKVSASHL----- 252
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 EVG-CWGEAHNCORFSLKNCSPQCSQGRCFGPK-----RECCHFCAGG 209
QY 253 ---EKHSDCLACLHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 309
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 CTGPTQSDCLACKNFYDGVCKQECPPMQIYNPTNYFWPNPDGKYATGATCRKCP-EH 268
QY 310 LSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVTSANIOE 369
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 LLKDNAGACVRKCPKGMQNSE-----CVPCKGVCPKTCPEGIVH-----SDNIGN 315
QY 370 FAGCKKIFGSLAFLPESFDGDPASNT-----APLQPEQLQVFTLEITGYLYISAW 421
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 YKDCITIEGSLDQSFDFGQQVYTNFSFGPRYIKIDPDRLEVFSTVKETGFINIOAH 375
```

```
QY 422 POSLPDLSVFQNLQVIRGRILHNGAY-SLTLOGLIGISWLGRLSLRELGLSLALIHNTHL 480
Db 376 HPNFTTLNFRNLVVGROLENKLPASVYIVKTSLSLEKLSLKRVSIGSIVILENSDL 435
QY 481 CFVHTVPDOLFRNPQHALLHTANRDECEVGEGLACHQLCARGHCWGPFGPTQVCNCSQF 540
Db 436 CFVEDIDWSEIKKSSDHEVMVQNRNATECHEEGMECSEQCSKAGCWGKGPEQCLECKNV 495
QY 541 LRQECVEECRVLQGLPREV-VNARHCLPCHPCOPONGSVTCFGEADOCVACAHYKOP 599
Db 496 KYKGRCLDSCK---SLPRLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSCMNVKOG 547
QY 600 PFCVARECP-----ACQPCPINCT-----HSCVDL-----DD-----KG 647
Db 548 RFCVAECPTTKHAMNGTCINCHKTCVCGRPRDTIAPDGCISCDKALIGSDAKIERCLMK 607
QY 619 INKFPD-----REG-----RASPLTSIVSAVVGILLVVLGVVFGI--- 679
Db 608 DESCPOGYVYDYVLQEBEPLKQLSGKAVCRKCHPRCKKCTGYGFHEQFCOECTGYKKGEQ 667
QY 628 -----ACQPCPINCT-----HSCVDL-----DD-----KG 647
Db 668 CEDECPODFYANEETRICLPCHQECRCGCHGLGDDHCECNLKLFEQDPYDNATFTCVSN 727
QY 648 CPAEQ-----RASPLTSIVSAVVGILLVVLGVVFGI--- 679
Db 728 CPASHVPYKRPQEBAGKIGPYCSADSMQSLRIEPTQVKIWMGSMVALILLCVVFGIAFV 787
QY 680 LKRRQOKRKYTMRLLOTELVELPTPSGAMPNQAOMRIKTELKVKVVLGSGAFGT 739
Db 788 LFSRHKNKDAVAKNTWALACEDSEPLRPSNVGNPLTKLRIIEASIRRGVGLMGAFGR 847
QY 740 VYGIWIPGGENVKIPVAIKVLRENTSPKANKILDEAYVMAGVGPYSRLLGICLTST 799
Db 848 VFKGVMPGESVKIPVAIKVLMEMSGSESKFLEEAYIMASVEHPNLLKLLAVCMTSQ 907
QY 800 VOLVQLMPYGCLLDHVRENRLGSGODLLNWCQIAKMSYLEDVRLVHRDLAARNVLV 859
Db 908 MMLITQLMPLGCLLDVYRNKDKIGSKALLNWSQTQARGMAYLEERLVRHDLAARNVLV 967
QY 860 KSPNVHKITDFGLARLLIDETEHADGKGVPIKWALESILRRRTHQSDVWSYGVTVW 919
Db 968 QTPSCVKITVFGAKLLDFDSDEYRAAGGKMPKWLALCEIRHRVFTSKSDVMWAFGITW 1027
QY 920 ELMTFCAKPYDGIIPAREIPDLLEKGERLPPOPPICTIDVYMWKMWIDSECRPRFELV 979
Db 1028 ELLTYGARPYENVYPAKDVELIIEGHKLPOPDICSLDVCILLSCWVLDADARPTFKLA 1087
QY 980 SEFSRMARDPQRFVWQNEIDLGPASPLDSTFYRSLLLEDDDMGDLV----- 1024
Db 1088 ETPAEKARDPGRYLM-----PGDKFMRLPSYTNQDEKDLIRTLAPVMAAAAAA 1138
QY 1025 -----DAEEYLVPOQGFPCPDPAAGGWMVHRRSSSTRSGGDLTLGLE 1070
Db 1139 AGASNVDPSTIAETDEYLOPKTRPSIMLPGPSA-----VE 1175
QY 1071 PSEEEAPRS-----PLAP---SEGAGSDVFDGLGMGAAGKQLSLPHTHDPSPQRYSED 1121
Db 1176 PS-DEMPKSLRYCKDPLKPDDETGDHKEV-----GVGGIR----- 1210
QY 1122 PTVPLPSETDGYVAPITCSPQRYVNPQDVROPSPREGPLPAARPAGATLERAKTLSP 1181
Db 1211 --LNLPLDEDDYLMP-TCOSQ---NQS----- 1233
QY 1182 GRKNGVKDVFAFGAVENBEYL-----TPQGGNAOPHPHPAFSPAFDNLVYWDQD 1232
Db 1234 G-----YMDLIGVPASVDNBEYLMGMSQTQIAGLAQSGMG---PHTPP----- 1272
QY 1233 PPERGAPPSTFKGTPTAENPE 1253
Db 1273 -----PPNTPNGMPTHQHSQ 1287
```

```
RESULT 8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Evans A.J., Henner W.D.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 25.7%; Score 1752.5; DB 4; Length 419;
Best Local Similarity 83.1%; Pred. No. 3.2e-124;
Matches 339; Conservative 6; Mismatches 34; Indels 29; Gaps 4;

QY 1 MELAALCRMGILLALLPPGAASQVCTGDMKRLRPASPETHLMLRLHLHQGVVQGNL 60
Db 1 MELAALCRMGILLALLPPGAASQVCTGDMKRLRPASPETHLMLRLHLHQGVVQGNL 60
QY 61 ELTYLPTNASLSPQDIQEVQGVLIHQNQVPLORLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSPQDIQEVQGVLIHQNQVPLORLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGRLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNNTPTVTGASPGRLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNOLA 180
QY 181 LTLIDTNRGRACHPCSPMCKGSRGCGESSEDCQSLTRTVCCAGCARCKGPLPFTVSF 240
Db 181 LTLIDTNRGRACHPCSPMCKGSRGCGESSEDCQSLTRTVCCAGCARCKGPLPFTVSF 240
QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRVTFGASCVTACPNYVLTSDVGSCTLVCPHFNQEVTAEDGTORCEKCKKPCARVCYGLG 352
Db 287 GRVTFGASCVTACPNYVLTSDVGSCTLVCPHFNQEVTAEDGTORCEKCKKPCARVCYGLG 341
QY 353 MEHLREVRAVTSANIOEFAGCKKIFGSLAPFESFDGDPASNTAPLQ 400
Db 342 THSLPRPRAVPVPLRMQPG--PAHPVLSFLRPSMDLVSAFYSLPLAP 387

RESULT 9
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ
  Query Match      25.5%; Score 1739; DB 11; Length 367;
  Best Local Similarity 88.0%; Pred. No. 2.7e-123;
  Matches 333; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 954
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

QY 955 IDVYIMVCKWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1014
DB 61 IDVYIMVCKWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPSSPMDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOGGFCPPAPGACGMVHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEYLVPOGGFCPPALGTGSTAHRHRSSARSGGELTLGLEPSEE 180

QY 1075 EAPRSPAPSEAGSDVDFDGLGMAAKGLQSLTHDPSPLOQRYSDPTVPLPSETDGYV 1134
DB 181 EAPRSPAPSEAGSDVDFDGLAVGVTKGLQSLSPHDLSPLOQRYSDPTLPLPETDGYV 240

QY 1135 APLTCSPOEYVQNPQVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKVFAFG 1194
DB 241 APLTCSPOEYVQNPQVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKVFAFG 300

QY 1195 GAVENPEYLTPOGGAAPQHPAPFAPFADNLYWQDPPERCAPSTFKGTPTAENPEY 1254
DB 301 GAVENPEYLTPOGGAAPQHPAPFAPFADNLYWQDPPERCAPSTFKGTPTAENPEY 360

QY 1255 LGLDVPV 1261
DB 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ
  Query Match      25.2%; Score 1720; DB 15; Length 729;
  Best Local Similarity 54.8%; Pred. No. 2.1e-121;
  Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PQNGSVTCFGEADQCVACAHYKDPFCVARCSPGVKPDLSYMPIKVPDEEGACQCPPI 634
DB 141 PEETATPTKGP--DHCCKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHP 197

QY 635 NCTHSCVDLDDKGCRAEQRASPLTSTVSAVV--GILLVVLGVVFGVILIKERQOKIRKVTM 693
DB 198 NCTRCCKGCGLEGGCP---NGSKTPSTAAGVVGGLLCLLVVGLGIGLYLRRR-HIVRKRTL 253

QY 694 RRLQETELVEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVK 753
DB 254 RRLQERELVEPLTPSGEAPNQAHLRLAKETEFKVKVLGSGAGFTVYKGLWIPGEKVK 313

QY 754 IPVAIKVIRENTSPKANKEILDEAYVMAGVSPYSRLLGICLTSTVQLVTQMPYGCLL 813
DB 314 IPVAIKELREATSPKANKEILDEAYVMASVNDPRVCRLLGICLTSTVQLITQMPYGCLL 373

QY 814 DHVRENRLGSGDQLLNCWQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDQGLA 873
DB 374 DYIREHKONIGSQYLLNWCQIAKMNYLEERLVRDLAARNVLVKTPOHVKITDFGLA 433

QY 874 RLDDIDETEHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 933
DB 434 KLGDADEKEYHAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 493

QY 934 AREIPDLLEKGERLPQPPICTIDVYIMVCKWIDSECRPRFRELVSFEFSRMARDPQRFV 993
DB 494 ASEISVLEKGERLPQPPICTIDVYIMVCKWIDADSRPKFRELIAEFSKMARDPDPRYL 553

QY 994 VIQ-NEDLGPASPLDSTFYRSLDDMGDLVDAEYLVPOGGFCPPAPGACGMVHHR 1052
DB 554 VIQGERMHLPTDTSKFTYRLMEEDMEDIVDADEYLVPHQGGF----- 598

QY 1053 HRSSSTRSGGDLTLGLEPSEEBEAPRSP-----APSEAGSDVDFDGLGMAAKGLQSL 1107
DB 599 -NSPST-----SRTPLLSLSATSNSATNCID-----RNCQGH 631

QY 1108 PTHDPSPLORYSEDPTVPLPSET--DGYVAPLTCSPQEYVNPQVVRPQPSREGPLPA 1165
DB 632 PVREDSFVQRYSSDPTGNFLESIDDGFL-----PAPEYVQ--LMPKKPS----- 675

QY 1166 ARPAGATLERAKTILSPGKNGVVKVDF-----AFGGAVENPEYL 1203
DB 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
```

RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 2e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCACAHYKDPFCVARGPVGKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 643
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCRTGCKGP 57

QY 644 DDKGCPEAQRASPLTSIVSAV-GILLVVLGVVFGILLIKERQOKIRKYMRLLQETEL 702
Db 58 GLEGGP---NGSKTPSAAAGVGGGLVCLVVGIGLGLYLR--HIVKRTLRLLQEREL 113

QY 703 VEPLTSGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 762
Db 114 VEPLTSGEAPNOAHLILKETEFKVKVLGSGAGFTVYKGLIPEGEKVKIPVAIKELR 173

QY 763 ENTSPKANKEILDEAYVMAGVGPYSRLLGICLTSTVQLVTLQMPYGCILLDHVRENRR 822
Db 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLTQMPYGCILLDVIREHKN 233

QY 823 LGSQDLLNWCQIAKMSYLEDLVHRDLAARNVLKSPNHVKITDFGLARLLDDETE 882
Db 234 IGSOYLLNWCQIAKGNLYLEERLVRDLAARNVLKTPQHKITDFGLAKLLGADKE 293

QY 883 YHADGGKVPKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLE 942
Db 294 YHAEGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353

QY 943 KGERLPQPPCTIDVYIMVKWMDSECRPRFRELVSFMRARDPQRFVVIQ-NEDLG 1001
Db 354 KGERLPQPPCTIDVYIMVKWMDADSRPKFRELIAEFSKWARDPPRYLVIOGDERMH 413

QY 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSG 1061
Db 414 LPSPDTUSKFRTLWEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1062 GGDLTGLLEPSEEAAPRSL-----APSEGAGSDVFDGLGMAAKGLQLSLPHDPSPLQ 1116
Db 455 -----SRTPLLSLSLATSNNATNCID-----RNGQGHVREDSFVQ 491

QY 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPEYVNPQVVRPQPPSPREGPLPAARAGATLE 1174
Db 492 RYSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS----- 526

QY 1175 RAKTLSPGKNGVKVDF-----AFGGAVENPEYL 1203
Db 527 ----TAMVQNIYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PF1659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF18349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00108; EF HAND; UNKNOWN 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 4.5e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGERLPQPPICT 954
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGERLPQPPICT 60

QY 955 IDVYIMVKWMDISECRPRFRELVSFMRARDPQRFVVIQEDLGASPDLSTFYRSL 1014
Db 61 IDVYIMVKWMDISECRPRFRELVSFMRARDPQRFVVIQEDLGASPDLSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSGGDLTLGLPSEE 1074
Db 121 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSGGDLTLGLPSEE 180

QY 1075 EAPRSLAPSEGAGSDVFDGLGMAAKGLQLSLPHDPSPLQRYSEDPVPLPSETDGYV 1134
Db 181 EAPRSLAPSEGAGSDVFDGLGMAAKGLQLSLPHDPSPLQRYSEDPVPLPSETDGYV 240

QY 1135 APLTCSPOPEYVNPQVVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVKVDFVAFG 1194
Db 241 APLTCSPOPEYVNPQVVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVKVDFVAFG 300

QY 1195 GAVENPEYLTPOGGAAPQP-----HPPPA---FSPAFDNL 1226
Db 301 GAVENPEYLTPOGGAALSPTLLPSAQPSSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YYWD-QDPPER-----GAPPTFKGTPTAEN 1251
Db 361 WWVTCQPEPQVRRSPDVSSSGSREGLTSAGIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.

Query Match		24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity		54.9%; Pred. No. 6.5e-116;
Matches		345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;
QY	584	GPEADQCVACAHYKDPFCVACRCPGKPDLSYMPIWKPDEBEGACQCPINCTHSCVDL 643
DB	1	GP--DHCMKAFIDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 57
QY	644	DKGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMERLLOETEL 702
DB	58	GLEGCP---NGSKTFSIAGVGGLLCLVVGGLGGLYLRRL-HIVRKTLRLRLQEREL 113
QY	703	VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGTVYKIGWIPDGENVKIPVAIKVL 762
DB	114	VEPLTPSGEAPNOAHLRLKETELFEKVKVVLGFGAGTVYKGLWIPBEGKVTPVAIKEL 173
QY	763	ENTSPKANKEIIDEAYVMAGVSPVSRLLGLICLTSTVOLVTLQMPYGLLDHVRNRGR 822
DB	174	EATSPKANKEIIDEAYVMASVDNPHVCRLLGLICLTSTVOLVTLQMPYGLLDYIREHKN 233
QY	823	LGSQDLNMCMOIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDDTE 882
DB	234	IGSQDLNMCVQIAKGMNLEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
QY	883	YHADGGKVPKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLE 942
DB	294	YHAEGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGPASEISSVLE 353
QY	943	KGERLPQPPICITIDVYIMVKWCMIDSECRPFRELVSFMSRMADPQRFVVIQ-NEDIG 1001
DB	354	KGERLPQPPICITIDVYIMVKWCMDSADSRKPRFELIAEFMSRMADPQRYLVIQGDERMH 413
QY	1002	PASPLDSTFYRLLEDDDDGLVDAEYLVPOQGFPCDPAPGAGGVVHHRSSGSTRSG 1061
DB	414	LPSPTDSKPYRLMEEDMEDVDAEYLVPHQGGF-----NSPST----- 454
QY	1062	CGDLTLGLEPSEEEAPRSL-----APSEGAGSDVDFDGLGMGAAGKLOSLPTHDPSPLO 1116
DB	455	-----SRTLPLLSLTSATSNATNCIDRNG-----H----- 481
QY	1117	RYSEDPVPLPSETDGYVAPLTCSPQEPYVQDVRPQPPSPREGPLPAARPAAGAT-LER 1175
DB	482	-----PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQIQIYVISTATISK 523
QY	1176	AKTLPNGKGVKDVFAFGGAVENPEYL 1203
DB	524	LPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15

Q9WVF5	PRELIMINARY;	PRT;	655 AA.
AC	Q9WVF5;		
DT	01-NOV-1999 (Tremblrel. 12, Created)		
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).		
DE	isoform 3).		
GN	EGFR.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=LIVER;		
RA	Reiter J.L., Threadgill D.W., Danieleen A.J., Schell C.,		
RA	Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,		
RA	Maihle N.J.;		
RT	"Alternative Transcripts from the Human and Mouse EGFR Genes Encode		
RT	Carboxy-Terminal Truncated Receptors.";		
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RN	[2]		

RP	SEQUENCE FROM N.A.
RC	STRAIN=C3H/101.129/SVJ, AND 129/SVEVITAC;
RA	Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danieleen A.J.,
RA	Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA	Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA	Maihle N.J.;
RT	"Comparative genomic sequence analysis and isolation of human and
RT	mouse alternative Egr transcripts encoding truncated receptor
RT	isoforms.";
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6J; TISSUE=LIVER;
RC	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA	Hayashizaki Y.;
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690 (2001).
DR	EMBL; AF124513; AAD44149.1; -
DR	EMBL; AF275366; AAG28047.1; -
DR	EMBL; AF275364; AAG28047.1; JOINED.
DR	EMBL; AF275365; AAG28047.1; JOINED.
DR	EMBL; AK004944; BAB23688.1; -
DR	EMBL; AK004883; BAB23641.1; -
DR	EMBL; AK004911; BAB23662.1; -
DR	MGD; MGI:95294; Egfr.
DR	InterPro; IPR000494; EGFR_L domain.
-3	InterPro; IPR002174; Furin-like.
DR	Pfam; PF00757; Furin-like; 1.
DR	Pfam; PF01030; Recep_L domain; 2.
DR	SMART; SM00261; FU; 3.
KW	Receptor.
SQ	SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

FT Region /label= Cysteine_rich_domain
210..224
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
250..264
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
324..483
/label= Ligand_binding_domain
FT Region
325..339
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
369..383
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
465..479
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
484..623
/label= Cysteine_rich_domain
FT Region
579..593
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
624..654
/label= Transmembrane_domain
FT Region
632..652
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
653..667
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
655..1010
/label= Tyrosine_kinase_domain
FT Region
661..675
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
695..709
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Region
710..730
/label= insertion_region
/note= "suitable for foreign epitope insertion"
FT Domain
1011..1235
/label= C-terminal_domain
XX
WO200020027-A2.
XX
13-APR-2000.
XX
05-OCT-1999; 99WO-DK00525.
XX
05-OCT-1998; 98DK-0001261.
XX
20-OCT-1998; 98US-0105011.
XX
(MEBI-) M & E BIOTECH AS.
XX
Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
PI Gautam A, Birk P, Karlsson G;
XX
WPI; 2000-349917/30.
XX
N-FSDB; AAA09455.
XX
Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
Claim 62; Page 193-198; 220pp; English.
XX
This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC response. Subdominant CTL epitopes, antibody binding regions and
CC cysteine residues involved in disulfide bonds are preserved in the
CC immunogenized forms. Regions suitable for the insertion of foreign T
CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer
CC when the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6690; DB 21; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQGCQVVGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHYQGCQVVGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPTGASPGGLREQLRSITELKGGVLQIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTGASPGGLREQLRSITELKGGVLQIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLDTNRSRACHPCSPMCKSGRCWSESSEDCQSITRTVCAGGCARCKGPLPNNFTVSF 240
Db 181 LTLDTNRSRACHPCSPMCKSGRCWSESSEDCQSITRTVCAGGCARCKGPLPNNFTVSF 240
Qy 241 WLRVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 286
Qy 293 GRYTFGASCVTACPNYILSTDVGSCTLVCPHQNQVTAEDGTQRCSEKSKPCARVCYGLG 352
Db 287 GRYTFGASCVTACPNYILSTDVGSCTLVCPHQNQVTAEDGTQRCSEKSKPCARVCYGLG 346
Qy 353 MEHLREVRVNTSANTQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEEI 412
Db 347 MEHLREVRVNTSANTQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEEI 406
Qy 413 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSLRELGSGLA 466
Qy 473 LIHNTHLCFVHTVPWDQLFNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPWDQLFNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCYA 592
Db 527 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCYA 586
Qy 593 CAHYKDPFPFCVARCPGKVPDLSYMPWKFPDEEGACQPCPINCTHSCVDDDDKGCAPAQ 652
Db 587 CAHYKDPFPFCVARCPGKVPDLSYMPWKFPDEEGACQPCPINCTHSCVDDDDKGCAPAQ 646
Qy 653 RASPLTSIVSAVGLLLVVLGVVFGILIKRQOKIRKYTWRRLLQETLEVEPLTPSGAM 712
Db 647 RASPLTSIVSAVGLLLVVLGVVFGILIKRQOKIRKYTWRRLLQETLEVEPLTPSGAM 706
Qy 713 PNOQMRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKE 772
Db 707 PNOQMRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKE 766

QY 773 ILDEAYVMAGVGSPPVSRLLGICLTSTVQLVTOLMPYGCCLLDHVRNRLGSOILLNWC 832
 DB 767 ILDEAYVMAGVGSPPVSRLLGICLTSTVQLVTOLMPYGCCLLDHVRNRLGSOILLNWC 826
 QY 833 MQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVP 892
 DB 827 MQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVP 886
 QY 893 KWMALISILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
 DB 887 KWMALISILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 946
 QY 953 CTIDVYIMVKWCMIDSECRPRFRELVSERFMRDPORFVVIQNEDELGPASPLDSTFYR 1012
 DB 947 CTIDVYIMVKWCMIDSECRPRFRELVSERFMRDPORFVVIQNEDELGPASPLDSTFYR 1006
 QY 1013 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPS 1072
 DB 1007 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPS 1066
 QY 1073 EEEAPRSLAPSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSEDPTVPPLSETDG 1132
 DB 1067 EEEAPRSLAPSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSEDPTVPPLSETDG 1126
 QY 1133 YVAPLTCSPOEYVNPDPVRPQPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVA 1192
 DB 1127 YVAPLTCSPOEYVNPDPVRPQPPSPREGPLPAAPAGATLERAKTILSPGKNGVVKDVA 1186
 QY 1193 FGGAVENPEYLTPOGGAAPQHPHPAPSPFDNLYYWDQDPPERGAPPSTFKGTPTAENP 1252
 DB 1187 FGGAVENPEYLTPOGGAAPQHPHPAPSPFDNLYYWDQDPPERGAPPSTFKGTPTAENP 1246
 QY 1253 EYGLGDVPV 1261
 DB 1247 EYGLGDVPV 1255

RESULT 2

AAE12130
 ID AAE12130 standard; Protein; 1255 AA.

AC AAE12130;

DT 18-DEC-2001 (first entry)

XX Human tyrosine kinase-type receptor, HER-2.

XX Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Region 774..782
 FT /note= "Antigenic epitope"

XX WO200168677-A2.

XX 20-SEP-2001.

XX 16-MAR-2001; 2001WO-US40328.

XX 16-MAR-2000; 2000US-0527487.

XX (GENZ) GENZYME CORP.

XX Nicolette CA;

XX WPI; 2001-616284/71.

XX N-PSDB; AAD19731.

XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.

XX The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

XX Sequence 1255 AA;

Query Match 98.0%; Score 6690; DB 22; Length 1255;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASTOVCTGDMKLLPASPETHLDMLRLHYGCCVVOGNL 60
 DB 1 MELAALCRWGLLALLPPGAASTOVCTGDMKLLPASPETHLDMLRLHYGCCVVOGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNVQVQVLPQRLRVIRGTQLPEDNVALAVLNG 120
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNVQVQVLPQRLRVIRGTQLPEDNVALAVLNG 120
 QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGCARCKGPIPFNNFTVSF 240
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVGAGGCARCKGPIPFNNFTVSF 240
 QY 241 WLRVPKVSASHLE-----KHSDCLACILFHNHSGICELHCPALVTYNTDTFESMPNPE 292
 DB 233 -----TDCHEQCAAGCTGPKHSDCLACILFHNHSGICELHCPALVTYNTDTFESMPNPE 286
 QY 293 GRVTFGASCVTACPNYLSLTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
 DB 287 GRVTFGASCVTACPNYLSLTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
 QY 353 MEHLREVRATSANIQEFAGCKKIFGSLAFPESEFDGDPASNTAPLQEQLOVFTLBEI 412
 DB 347 MEHLREVRATSANIQEFAGCKKIFGSLAFPESEFDGDPASNTAPLQEQLOVFTLBEI 406
 QY 413 TGYLYISAWPDSLPLDLVFNQNLQVIRGRIILHNGAYSLSLTQGLGISWLGRLSRLGSLGA 472
 DB 407 TGYLYISAWPDSLPLDLVFNQNLQVIRGRIILHNGAYSLSLTQGLGISWLGRLSRLGSLGA 466
 QY 473 LIHNHNLCLFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAGCHQICARGHCWGPGPT 532
 DB 467 LIHNHNLCLFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLAGCHQICARGHCWGPGPT 526
 QY 533 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQVCA 592
 DB 527 QCVNCSQFLRGQECVBEICRVLQGLPRBYVNAHCLPCHPECPQNGSVTCFPGBADQVCA 586
 QY 593 CAHYKDPFPFCVARCPSGVKPDLSYMPWKPEDEEGACQPCINCTHSCVDLDDKGCRAEQ 652
 DB 587 CAHYKDPFPFCVARCPSGVKPDLSYMPWKPEDEEGACQPCINCTHSCVDLDDKGCRAEQ 646

Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db |||||
Qy 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Db |||||
Qy 713 PNOAQMRILKETELURKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 772
Db |||||
Qy 707 PNOAQMRILKETELURKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 766
Db |||||
Qy 773 ILDEAYVMAGVSPVSRLLGICITSTVQLVTLQMPYGCILLDHDVRENRRGLSQDILLNWC 832
Db |||||
Qy 767 ILDEAYVMAGVSPVSRLLGICITSTVQLVTLQMPYGCILLDHDVRENRRGLSQDILLNWC 826
Db |||||
Qy 833 MQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPPI 892
Db |||||
Qy 827 MQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPPI 886
Db |||||
Qy 893 KWMALLESILRRRFTHQSDVWVSGYVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPPI 952
Db |||||
Qy 887 KWMALLESILRRRFTHQSDVWVSGYVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPPI 946
Db |||||
Qy 953 CTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db |||||
Qy 947 CTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
Db |||||
Qy 1013 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLEPS 1072
Db |||||
Qy 1007 SLEDDDDMGDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLEPS 1066
Db |||||
Qy 1073 EEEAPRSLAPSEGAGSDVFGDGLGMAKGLQSLPHDPSFLQRYSEDPTVPLPSETDG 1132
Db |||||
Qy 1067 EEEAPRSLAPSEGAGSDVFGDGLGMAKGLQSLPHDPSFLQRYSEDPTVPLPSETDG 1126
Db |||||
Qy 1133 YVAPLTCSPQPEYVYNQDVRPQPPSPREGPLPAARPAATILERAKTLSPGKNGVVKDVF 1192
Db |||||
Qy 1127 YVAPLTCSPQPEYVYNQDVRPQPPSPREGPLPAARPAATILERAKTLSPGKNGVVKDVF 1186
Db |||||
Qy 1193 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQPPERGAPPSTFKGTPTAENP 1252
Db |||||
Qy 1187 FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQPPERGAPPSTFKGTPTAENP 1246
Db |||||
Qy 1253 EYLGLDVVP 1261 /
Db |||||
Qy 1247 EYLGLDVVP 1255
Db |||||
RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX Homo sapiens.
OS Synthetic.
OS
PN WO200100244-A2.
XX
XX 04-JAN-2001.
PD
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
XX (GETH) GENENTECH INC.
XX
PI Erickson S, Schwalli R;

XX WPI; 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6690; DB 22; Length 1255;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHDMLRHLGYCQVVGNL 60
Db |||||
Qy 1 MELAALCRWGLLLALLPPGAASCTVCTGDMKRLPASPETHDMLRHLGYCQVVGNL 60
Db |||||
Qy 61 ELTYLPTNASLFLQDIOEVQYVLIHNOVROVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db |||||
Qy 61 ELTYLPTNASLFLQDIOEVQYVLIHNOVROVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db |||||
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSILTEILKGVLIQRPOLCYODTILKWDIFHKNQLA 180
Db |||||
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSILTEILKGVLIQRPOLCYODTILKWDIFHKNQLA 180
Db |||||
Qy 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACAGGCARCKGPIPNFTVSF 240
Db |||||
Qy 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACAGGCARCKGPIPNFTVSF 232
Db |||||
Qy 241 WLRVPKVSASHLE-----KHSDDLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db |||||
Qy 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
Db |||||
Qy 293 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db |||||
Qy 287 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
Db |||||
Qy 353 MEHLREVRVTSANIQBFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEBEI 412
Db |||||
Qy 347 MEHLREVRVTSANIQBFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEBEI 406
Db |||||
Qy 413 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSLRELSGLA 472
Db |||||
Qy 407 TGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSLRELSGLA 466
Db |||||
Qy 473 LIHNTHTLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQOLCARGHCHGPGPT 532
Db |||||
Qy 467 LIHNTHTLCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQOLCARGHCHGPGPT 526
Db |||||
Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 592
Db |||||
Qy 527 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 586
Db |||||
Qy 593 CAHYKDPFPFCVARCPGVKPDLSYMPWKFPDEGACQPCPINCTHSCVDLDDKGCBAEQ 652
Db |||||
Qy 587 CAHYKDPFPFCVARCPGVKPDLSYMPWKFPDEGACQPCPINCTHSCVDLDDKGCBAEQ 646
Db |||||
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db |||||
Qy 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Db |||||
Qy 713 PNOAQMRILKETELURKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 772
Db |||||

Db 707 PNOAQRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKE 766
 Qy 773 ILDEAYVMAGVSPVYSRLGICLTSTVQLTQMPYGCCLLDHVRNRRGRLGSODLLNWC 832
 Db 767 ILDEAYVMAGVSPVYSRLGICLTSTVQLTQMPYGCCLLDHVRNRRGRLGSODLLNWC 826
 Qy 833 MQIAKGSYLEDVRLVHRDLAARNVVKSPHVKITDFGLARLLDIDETEHADGGKVP 892
 Db 827 MQIAKGSYLEDVRLVHRDLAARNVVKSPHVKITDFGLARLLDIDETEHADGGKVP 886
 Qy 893 KMALESILRRFTHOSDVMWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 952
 Db 887 KMALESILRRFTHOSDVMWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 946
 Qy 953 CTIDVYIMVWKWIDSECRPRFRELVSFSEFARMARDPQRFVVYIQNEDLGASPLDSTFYR 1012
 Db 947 CTIDVYIMVWKWIDSECRPRFRELVSFSEFARMARDPQRFVVYIQNEDLGASPLDSTFYR 1006
 Qy 1013 SLEDDDMGDLVDAEYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLBPS 1072
 Db 1007 SLEDDDMGDLVDAEYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLBPS 1066
 Qy 1073 EEEAPRSLAPSEAGSDVDFQDGLCMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDG 1132
 Db 1067 EEEAPRSLAPSEAGSDVDFQDGLCMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDG 1126
 Qy 1133 YVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFA 1192
 Db 1127 YVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVDVFA 1186
 Qy 1193 FGGAVENPEYLTPOGGAAPQHPHPPAFSPADNLYWDDPBERGAPPSTFKGTPTAENP 1252
 Db 1187 FGGAVENPEYLTPOGGAAPQHPHPPAFSPADNLYWDDPBERGAPPSTFKGTPTAENP 1246
 Qy 1253 EYLGLDVVPV 1261
 Db 1247 EYLGLDVVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX Human HER2 (Erbb2) polypeptide.
 DE
 XX Human; HER2; Erbb; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; Erbb2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 XX US2002001587-A1.
 PN
 XX 03-JAN-2002.
 PD
 XX 16-MAR-2001; 2001US-081123.
 PF
 XX 16-MAR-2000; 2000US-189844P.
 PR
 PR 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWKOWSKI M.
 XX

PI Erickson S, Schwall R, Sliwkowski M;
 XX WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, Erbb or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (Erbb) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses Erbb2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (Erbb2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.0%; Score 6690; DB 23; Length 1255;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1241; Conservative 1; Mismatches 5; Indels 22; Gaps 2;
 Qy 1 MELAALCRWGLLLALLPFGAASSTOCTGCTDMKRLRPASPETHLDMRLHYQCQVVQGNL 60
 Db 1 MELAALCRWGLLLALLPFGAASSTOCTGCTDMKRLRPASPETHLDMRLHYQCQVVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIEQVQGYVLIHQNVRQVPLQRLRVRGTQFEDNYALAVLNG 120
 Db 61 ELTYLPTNASLSFLQDIEQVQGYVLIHQNVRQVPLQRLRVRGTQFEDNYALAVLNG 120
 Qy 121 DPLNNTPTVTCASPGGLREQLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
 Db 121 DPLNNTPTVTCASPGGLREQLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180
 Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCGPLPNNFTVSF 240
 Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCGPLP 232
 Qy 241 WLVRPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 292
 Db 233 -----TDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPE 286
 Qy 293 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
 Db 287 GRVTFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLG 346
 Qy 353 MEHLREVRVAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEE 412
 Db 347 MEHLREVRVAVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEE 406
 Qy 413 TGYLYISAWPDSLPLDSVFNQNLQVIRGRIHLHNGAYSLTQGLGISWGLRSLRELGSGLA 472
 Db 407 TGYLYISAWPDSLPLDSVFNQNLQVIRGRIHLHNGAYSLTQGLGISWGLRSLRELGSGLA 466
 Qy 473 LIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGGLACHQLCARGHCWGPPT 532
 Db 467 LIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVCGGLACHQLCARGHCWGPPT 526
 Qy 533 QCVNCSOFLRGQECVBEICRVLOGLPREVYNARHCLPCHPECPONGSVTCFGEADQVVA 592
 Db 527 QCVNCSOFLRGQECVBEICRVLOGLPREVYNARHCLPCHPECPONGSVTCFGEADQVVA 586
 Qy 593 CAHYKDPFPFCVARCPSGVKPDLSYMPYWKFPDEBAGCQPCPINCTHSCVDLDDKGCPEAQ 652


```
|||||
587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEACQPCPINCTSHSCVDLDDKCPAEQ 646
653 RASPLTSTVSAVGTLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAM 712
647 RASPLTSTVSAVGTLVVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAM 706
713 PNOQMIRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
707 PNOQMIRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 766
773 ILDEAYNAGVSGPVVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLGSDQLLNC 832
767 ILDEAYNAGVSGPVVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLGSDQLLNC 826
833 MOIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 892
827 MOIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 886
893 KWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
887 KWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 946
953 CTIDVYMIWVKWIMIDSCRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
947 CTIDVYMIWVKWIMIDSCRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
1013 SLEDDDDMGDLVDAEYLVPQGGFCPPAPAGAGMVRHRSSTRSGGDLTLGLEPS 1072
1007 SLEDDDDMGDLVDAEYLVPQGGFCPPAPAGAGMVRHRSSTRSGGDLTLGLEPS 1066
1073 EEEAPRSLAPFSEGAGSDVFDGDLGMAAGLQSLPTHDPSPQRYSEDPTVPLSETDG 1132
1067 EEEAPRSLAPFSEGAGSDVFDGDLGMAAGLQSLPTHDPSPQRYSEDPTVPLSETDG 1126
1133 YVAPLTCSPQRYNQPVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKQVFA 1192
1127 YVAPLTCSPQRYNQPVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKQVFA 1186
1193 FGGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPAENP 1252
1187 FGGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPAENP 1246
1253 EYLGLDVVP 1261
1247 EYLGLDVVP 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX AC AAW01111;
XX DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX KW breast cancer; ovary cancer; colon cancer; lung cancer;
XX KW prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key
XX FT Domain
XX FT 676...1255
XX FT /label= intracellular_domain
XX FT /note= "claimed domain, useful for immunisation"
XX PN W09630514-A1.
XX PD 03-OCT-1996.
XX
```

```
PF 28-MAR-1996; 96WO-US01689.
PR 31-MAR-1995; 95US-0414417.
PA (UNIW ) UNIV WASHINGTON.
XX Cheever MA, Disis ML;
XX WPI; 1996-455361/45.
XX N-PSDB; AAT40739.
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX Claim 2; Page 56-61; 71pp; English.
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
XX the product of the HER-2/neu oncogene (see also AAT40739). The
XX protein is over-expressed in various cancers, including breast,
XX ovarian, colon, lung and prostate. The intracellular domain of the
XX protein can be used to immunise an animal against a malignancy with
XX which the oncogene is associated. The polypeptide can be produced
XX in transformed host cells for use in immunisation. Alternatively,
XX animal cells are transfected in vivo or ex vivo with a viral vector
XX that directs expression of the polypeptide.
XX Sequence 1255 AA;
Query Match 97.9%; Score 6684; DB 17; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLLALLPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCVVQGNL 60
DB 1 MELAALCRWGLLLALLPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCVVQGNL 60
QY 61 ELTYLPTNASLFLQDIEQVGYVLIHNVQVQPLQRLRIVRGTLQFEDNVALAVLNG 120
DB 61 ELTYLPTNASLFLQDIEQVGYVLIHNVQVQPLQRLRIVRGTLQFEDNVALAVLNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGGCARCKGPLPFFNFTVSF 240
DB 181 LTLIDTNRSRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGGCARCKGPLPFFNFTVSF 240
QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
DB 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRYTTFGASCVTACPNYLSTDVGSCITLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
DB 287 GRYTTFGASCVTACPNYLSTDVGSCITLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
QY 353 MEHLREVRVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLBEI 412
DB 347 MEHLREVRVTSANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLBEI 406
QY 413 TGYLYISAWPDSLPLDSVFNQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLA 472
DB 407 TGYLYISAWPDSLPLDSVFNQVIRGRILHNGAYSLTLOGLGISWGLRSLRELGSGLA 466
QY 473 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCKWPGPT 532
DB 467 LIHNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECEVGEGLACHQLCARGHCKWPGPT 526
QY 533 QCVCNSQFLRGQECVBEICRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQVCA 592
DB 527 QCVCNSQFLRGQECVBEICRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQVCA 586
```



```
QY 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQ 652
DB 587 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQ 646
QY 653 RASPLTSIVSAVVGILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVPLTPSGAM 712
DB 647 RASPLTSIIISA VVGILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVPLTPSGAM 706
QY 713 PNOAQMRIILKTELRLKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
DB 707 PNOAQMRIILKTELRLKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
QY 773 ILDEAYVMAGVSPYVSRLLGLICTSTVOLVTQMLPYGCLLDHVRNRRGLSGDQLLNC 832
DB 767 ILDEAYVMAGVSPYVSRLLGLICTSTVOLVTQMLPYGCLLDHVRNRRGLSGDQLLNC 826
QY 833 MOIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 892
DB 827 MOIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 886
QY 893 KMALESILRRRFTHOSDVMYSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 952
DB 887 KMALESILRRRFTHOSDVMYSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP 946
QY 953 CTIDVYMIWVKCMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
DB 947 CTIDVYMIWVKCMIIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SILEDDMDGLVDAAEYLVPOQFFCPDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPS 1072
DB 1007 SILEDDMDGLVDAAEYLVPOQFFCPDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQLSLPTHDPSPQLQRYSEDTVPPLPSETDG 1132
DB 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQLSLPTHDPSPQLQRYSEDTVPPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVA 1192
DB 1127 YVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVA 1186
QY 1193 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENP 1252
DB 1187 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
DB 1247 EYLGLDVPV 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX
```

```
PA (CORI-) CORIXA CORP.
PA (SMIK) SMITHKLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX
DR WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
Claim 52; Fig 7; 128pp; English.
PS
XX The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 97.9%; Score 6684; DB 21; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRLPASPEHLDMRLHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRLPASPEHLDMRLHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNRQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNRQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSITELKGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSITELKGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180
QY 181 LTLIDNRSRACHPCSPCKGSRGSESDQSLTRTVACGACRCKPLPENNFTVSF 240
DB 181 LTLIDNRSRACHPCSPCKGSRGSESDQSLTRTVACGACRCKPLPENNFTVSF 240
QY 241 WLVRPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPE 292
DB 233 -----TDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPE 286
QY 293 GRYTEGASCVTACPNYLSLTDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLG 352
DB 287 GRYTEGASCVTACPNYLSLTDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLG 346
QY 353 MEHLREVRATVSANIQEFAGCKKIFGSLAFLPESDGDPSANTAPLQPSOLOVFETLEE 412
DB 347 MEHLREVRATVSANIQEFAGCKKIFGSLAFLPESDGDPSANTAPLQPSOLOVFETLEE 406
QY 413 TGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLSLTQGLGISWGLSRLSGSLA 472
DB 407 TGYLYISAWPDSLPLDSVFQNLQVIRGRILHNGAYSLSLTQGLGISWGLSRLSGSLA 466
QY 473 LIHNNTHLCFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 532
DB 467 LIHNNTHLCFVHTVPWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 526
QY 533 QCVNCSQFLRGQECVVEECRLVQLPREYVNAHRCLPCHPECPQNGSVTCFGEADQCV 592
DB 527 QCVNCSQFLRGQECVVEECRLVQLPREYVNAHRCLPCHPECPQNGSVTCFGEADQCV 586
QY 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQ 652
```

Db 587 CAHYKDPFCVACPSGVFDSL SYMPIWKFFDEGACQPCPINCTHSCVDLDDKGPAAEQ 646
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAM 712
Db 647 RASPLTSIISAVVGIILLVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAM 706
Qy 713 PNOAQRILKETELRKVKVLGSGAFQTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSGAFQTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSPYVSRLLGLICTSTVOLVTQLMYPGCLLDHVRNRLGSDLLNNC 832
Db 767 ILDEAYVMAGVSPYVSRLLGLICTSTVOLVTQLMYPGCLLDHVRNRLGSDLLNNC 826
Qy 833 MOIAKMSYLEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVP 892
Db 827 MOIAKMSYLEVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVP 886
Qy 893 KMALESILRRRFTTHQSDVMSYGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPDPI 952
Db 887 KMALESILRRRFTTHQSDVMSYGVTVWELMTCAKPYDGIIPAREIPDLLEKGERLPDPI 946
Qy 953 CTIDVYIMVCKMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVCKMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFYR 1006
Qy 1013 SLEDDDDMDGLVDABEYLVPOQGFCDPDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMDGLVDABEYLVPOQGFCDPDPAPGAGMVHRRSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQSLPHTDPSPLORYSEDPVLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQSLPHTDPSPLORYSEDPVLPSETDG 1126
Qy 1133 YVAPLTCSQPEYVNVQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVA 1192
Db 1127 YVAPLTCSQPEYVNVQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVA 1186
Qy 1193 FCGAVENBEYLTPOGGAAPQHPHPPAFSPAFONLYYWDODPPERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENBEYLTPOGGAAPQHPHPPAFSPAFONLYYWDODPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGDDVPV 1261
Db 1247 EYLGDDVPV 1255

RESULT 8

AAy84780

ID AAY84780 standard; Protein; 1255 AA.

XX AAY84780;

AC AAY84780;

DT 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbB-2 receptor protein.

DE SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;

KW tumor cell proliferation; tissue degeneration; arthropathy;

KW bone resorption; inflammatory disease; degenerative disorder;

XX wound healing.

OS Homo sapiens.

OS WO200020579-A1.

PN 13-APR-2000.

PD 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

PF 02-OCT-1998; 98US-0165192.

XX

XX

XX

XX

XX

XX

XX

XX

PA (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

PI WPI: 2000-303768/26.

DR N-PSDB; AAA14812.

XX

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE

PT erbB-2, inhibitors of the protein are useful for treatment of cancer -

PT Claim 3; Fig 2; 60pp; English.

PS

XX The present sequence represents a SPLICE erbB-2 receptor protein. The

XX protein has an in-frame deletion of 16 amino acids, 2 of which are

CC conserved cysteine residues, compared to the unspliced protein. The

CC erbB-2 polynucleotide is used to construct probes for detecting

CC disorders of cell transformation such as cancer. Antibodies to the

CC protein may be used to detect SPLICE erbB-2 in a sample. Agents

CC (e.g. antisense oligonucleotides) which inhibit the expression of

CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and

CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful

CC for treating conditions involving damaged cells including conditions

CC in which degeneration of tissue occurs, such as arthropathy, bone

CC resorption, inflammatory diseases, degenerative disorders of the

CC central nervous system and wound healing.

XX

SQ Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 21; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

Qy 1 MELAALCRWGLLALLPFGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCQVVGQNL 60

Db 1 MELAALCRWGLLALLPFGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCQVVGQNL 60

Qy 61 ELTYLPTNASLFLQDIOEVQGYVLIANHQRQVPLQRLRIRVGTQFEDNYALVDNG 120

Db 61 ELTYLPTNASLFLQDIOEVQGYVLIANHQRQVPLQRLRIRVGTQFEDNYALVDNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKQKOLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKQKOLA 180

Qy 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACRCKPLPNNFTVSF 240

Db 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACRCKPLPNNFTVSF 240

Qy 241 WLURVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292

Db 241 WLURVPKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292

Qy 293 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 352

Db 293 GRYTFGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLG 352

Qy 353 MEHLREVRATVSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQEPQLQVFTLESI 412

Db 353 MEHLREVRATVSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQEPQLQVFTLESI 412

Qy 413 TGLYLISAWPDSLPDLVSFQNLQVIRGRILHNGAYSLSLTQGLGISWGLRSURELSGLA 472

Db 413 TGLYLISAWPDSLPDLVSFQNLQVIRGRILHNGAYSLSLTQGLGISWGLRSURELSGLA 472

Qy 473 LIHNTHLFCFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCGPGPT 532

Db 473 LIHNTHLFCFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCGPGPT 532

Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFGEADQCV 592

Db 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTFCFGEADQCV 592

Qy 592 CAHYKDPFCVACPSGVKPDLSYMPIWKFFDEGACQPCPINCTHSCVDLDDKGPAAEQ 652

Db 592 CAHYKDPFCVACPSGVKPDLSYMPIWKFFDEGACQPCPINCTHSCVDLDDKGPAAEQ 652

Db 587 CAHYKDPFCVACPSGVKPDLSYMPIWKPPDEGACQPCPINCTHSCVDLDDKGCFAEQ 646
Qy 653 RASPLTSIVSAVVGILLVWLVGVVFGILIKRROOKIRKYTMRLLOETELVEPLTSGAM 712
Db 647 RASPLTSIIISAVVGILLVWLVGVVFGILIKRROOKIRKYTMRLLOETELVEPLTSGAM 706
Qy 713 PNOAQMRILKETELRKVKVGLSGAFGVTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQMRILKETELRKVKVGLSGAFGVTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAVYMACVGSPPYVSRLLGICLTSTVQLVTOIMPYGCULLDHVRENRLGSDLLNWC 832
Db 767 ILDEAVYMACVGSPPYVSRLLGICLTSTVQLVTOIMPYGCULLDHVRENRLGSDLLNWC 826
Qy 833 MQIAKMSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDDETYHADGGKVPI 892
Db 827 MQIAKMSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDDDETYHADGGKVPI 886
Qy 893 KMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 887 KMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVKCWMIDSECRPRFRELVSFERSMARDPQRFVVIQNEDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVKCWMIDSECRPRFRELVSFERSMARDPQRFVVIQNEDLGPASPLDSTFYR 1006
Qy 1013 SLEDDDDMGDLVDAEYLVPOQGFCCPDAPAGGMVHRRHRSSTSGGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLVPOQGFCCPDAPAGGMVHRRHRSSTSGGGDLTLGLEPS 1066
Qy 1073 EEBAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPVPLPSETDG 1132
Db 1067 EEBAPRSPLAPSGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPVPLPSETDG 1126
Qy 1133 YVAPLTCSPQPEYVNOQDVPRQPPSPREGPLPAARPAAGATLERPKTLSPKNGVVKDVFA 1192
Db 1127 YVAPLTCSPQPEYVNOQDVPRQPPSPREGPLPAARPAAGATLERPKTLSPKNGVVKDVFA 1186
Qy 1193 FGAVENPEYLTPOGGAAPOPHPPAPFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FGAVENPEYLTPOGGAAPOPHPPAPFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVVP 1261
Db 1247 EYLGLDVVP 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX AAB85458;
AC
XX
XX 25-SEP-2001 (first entry)
DT
DE Human HER-2/neu protein.
XX
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
XX Homo sapiens.
XX
XX WO200153463-A2.
XX
XX 26-JUL-2001.
XX
XX 19-JAN-2001; 2001WO-US01850.
PF
XX
XX 21-JAN-2000; 2000US-0177545.
PR
XX
XX (CORI-) CORIXA CORP.
PA
XX

PI
XX
DR
DR
XX
PT
PT
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
SQ

Cheever MA, Hand-Zimmermann S;
WPI; 2001-476112/51.
N-PSDB; AAH23392.
New antigen-presenting cells, useful as vaccines for eliciting or
enhancing an immune response to HER-2/neu protein, particularly useful
for treating or preventing cancer, e.g. breast cancer
Claim 2; Page 41-46; 49pp; English.

The invention provides an isolated antigen-presenting cell, which
expresses at least an immunogenic portion of a polypeptide that produces
an immune response to HER-2/neu protein. The antigen-presenting cells are
useful as vaccines for eliciting or enhancing an immune response to
HER-2/neu protein, particularly in treating or preventing malignancies in
which the HER-2/neu oncogene is associated. Specifically, these are
useful for treating or preventing cancer, e.g. breast cancer, ovarian,
colon, lung or prostate cancers. The present sequence represents
the human HER-2/neu protein (also known as p185 or c-erbB2).

Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 22; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLVYGCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLVYGCQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVVLAHNOVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNKOLA 180
Db 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNKOLA 180
Qy 181 LTLDITNRSRACHPCSPMKGSRGWGSESSDCSLARTVCAGGACRCKGPLPFNNFTVSF 240
Db 181 LTLDITNRSRACHPCSPMKGSRGWGSESSDCSLARTVCAGGACRCKGPLPFNNFTVSF 232
Qy 241 WLRVPKVSASHLE-----KHSDCLACILFHNHSGICELHCPALVYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACILFHNHSGICELHCPALVYNTDTFESMPNPE 286
Qy 293 GRYTFGASCVTACPNYLTVDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLG 352
Db 287 GRYTFGASCVTACPNYLTVDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARVCYGLG 346
Qy 353 MEHLREVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEI 412
Db 347 MEHLREVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEI 406
Qy 413 TGYLYISAWPDSLPLDSVFONQVIRGRILHNGAYSLLTQGLGISWGLSLRSLGSLA 472
Db 407 TGYLYISAWPDSLPLDSVFONQVIRGRILHNGAYSLLTQGLGISWGLSLRSLGSLA 466
Qy 473 LIHNTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGPT 532
Db 467 LIHNTHLCFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGPT 526
Qy 533 QCVNCSQFLRGQECVCECVRLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVA 592
Db 527 QCVNCSQFLRGQECVCECVRLQGLPREYVNAHCLPCHPECOQNGSVTCFGEADOCVA 586
Qy 593 CAHYKDPFCVACPSGVKPDLSYMPIWKPPDEGACQPCPINCTHSCVDLDDKGCFAEQ 652
Db 587 CAHYKDPFCVACPSGVKPDLSYMPIWKPPDEGACQPCPINCTHSCVDLDDKGCFAEQ 646
Qy 653 RASPLTSIVSAVVGILLVWLVGVVFGILIKRROOKIRKYTMRLLOETELVEPLTSGAM 712

```
Db 647 RASPLTSIIISAVGILLVVVLGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAM 706
Qy 713 PNOQMRLIKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 772
Db 707 PNOQMRLIKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSGPVSVRLGICLTSTVOLVTQLMPYGCCLLDHVRENRGLGSDLLNWC 832
Db 767 ILDEAYVMAGVSGPVSVRLGICLTSTVOLVTQLMPYGCCLLDHVRENRGLGSDLLNWC 826
Qy 833 MOIAGMSYLEDVRLVHDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKVP 892
Db 827 MOIAGMSYLEDVRLVHDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGKVP 886
Qy 893 KMALESILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 952
Db 887 KMALESILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 946
Qy 953 CTIDVYIMVWKWIDSECRPRFRELVSFMRMADPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVWKWIDSECRPRFRELVSFMRMADPQRFVVIQNEIDLGPASPLDSTFYR 1006
Qy 1013 SLLEDDDMGDLVDAAEYLVPOQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLRPS 1072
Db 1007 SLLEDDDMGDLVDAAEYLVPOQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLRPS 1066
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOKYSDDPTVPLPSETDG 1126
Qy 1133 YVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKD 1192
Db 1127 YVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARPAAGATLERAKTLSPGKNGVVKD 1186
Qy 1193 FGGAVENPEYLTPOGGAAPPHPPAFSPAFDNLVYWDQDPPERCAPPSTKGTPTAENP 1252
Db 1187 FGGAVENPEYLTPOGGAAPPHPPAFSPAFDNLVYWDQDPPERCAPPSTKGTPTAENP 1246
Qy 1253 EYLGDLVPV 1261
Db 1247 EYLGDLVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
AC AAG88267;
XX
XX
XX 11-SEP-2001 (first entry)
XX
XX HER2/neu amino acid sequence.
XX
XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
XX Homo sapiens.
XX
XX WO200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX PI Keogh E;
XX
XX WPI; 2001-374995/39.
```

```
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
PS Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
XX culture in vitro and binds to a complex of an epitope (I), bound to a
XX human leukocyte antigen (HLA) molecule; (2) a peptide (II), comprising (I)
XX and a second epitope and the peptide is less than 50 contiguous amino
XX acids that have 100% identity with a native peptide sequence of HER2/neu;
XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical
XX excipient; (4) an isolated nucleic acid encoding a peptide comprising
XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
XX and immunostimulant activities, and can be used in vaccines. (I), (II)
XX and (III) are useful for inducing cellular immune responses for the
XX prevention and treatment of cancer. (I) and (II) are useful for
XX monitoring or evaluating an immune response to a tumour-associated
XX antigen when incubated with a T lymphocyte sample from a patient and
XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope
XX based vaccines mean that immunosuppressive epitopes that may be present
XX in whole antigens may be avoided. Selected epitopes may be combined to
XX enhance immunogenicity. The possible pathological side effects caused by
XX infectious agents or whole protein antigen is eliminated. The vaccine
XX provides the ability to direct and focus an immune response to multiple
XX selected antigens from the same pathogen. Epitope-based anti-tumour
XX vaccines provides the opportunity to combine epitopes derived from
XX multiple tumour-associated molecules addressing the problem of tumour-
XX tumour variability and reducing the likelihood of tumour escape due to
XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
XX the exemplification of the present invention.
```

Seq Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 22; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

```
Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPETHLDMLRLHYQCCVQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPETHLDMLRLHYQCCVQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIEQVQGYVLIHNOVROVPLQRLRIVRGTLQEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIEQVQGYVLIHNOVROVPLQRLRIVRGTLQEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGRCWGSSEDCQSLTRTVCCAGCARCKGPLPFPNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGRCWGSSEDCQSLTRTVCCAGCARCKGPLPFPNNFTVSF 240
Qy 241 WLRPVKVSASHLE-----KHSDCLACLHFNHSGICELHCPALVTNTDFTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTNTDFTFESMPNPE 286
Qy 293 GRYTFGASCVTACPNYNYLSTDVSGCTLVCPILHNOEVTAEADGTQRCCKSKPCARVCYGLG 352
Db 287 GRYTFGASCVTACPNYNYLSTDVSGCTLVCPILHNOEVTAEADGTQRCCKSKPCARVCYGLG 346
Qy 353 MEHLREVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEEI 412
Db 347 MEHLREVRVTSANIQEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEEI 406
Qy 413 TGYLYISAWPDSLPLDSVFNQLQVIRGILHNGAYSILTQGLGTSWLGSLRELGSGLA 472
Db 407 TGYLYISAWPDSLPLDSVFNQLQVIRGILHNGAYSILTQGLGTSWLGSLRELGSGLA 466
Qy 473 LIHNHTHLCFVHTVPWDQLFRNPHQALHTANRPEDECVGGLGACHQLCARGHCWGPGPT 532
```

Db 467 LIHNTHLCFVHTVPMDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYNARHCLPCHPECOFQNGSVTCFGEADQOVA 592
Db 527 QCVNCSQFLRGQECVEECRVLQGLPREYNARHCLPCHPECOFQNGSVTCFGEADQOVA 586
Qy 593 CAHYKDPFPCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 652
Db 587 CAHYKDPFPCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 646
Qy 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVPLEPSPGAM 712
Db 647 RASPLTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVPLEPSPGAM 706
Qy 713 PNOAQMRILKTELRLKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKE 772
Db 707 PNOAQMRILKTELRLKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSPYVSRLLGLCTSTVOLVQLMPYGCCLLDHVRNRRGLSGDQLLNNC 832
Db 767 ILDEAYVMAGVSPYVSRLLGLCTSTVOLVQLMPYGCCLLDHVRNRRGLSGDQLLNNC 826
Qy 833 MOIAGMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPI 892
Db 827 MOIAGMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPI 886
Qy 893 KWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPI 952
Db 887 KWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPI 946
Qy 953 CTIDVYIMVWCWIMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVWCWIMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEDLGPASPLDSTFYR 1006
Qy 1013 SLEDDDDMDGLVDAEYLVPOQGFPCDPAPGAGGVHRRHSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMDGLVDAEYLVPOQGFPCDPAPGAGGVHRRHSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQLSPTHTDPSQLRYSIEDPTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAGKQLSPTHTDPSQLRYSIEDPTVPLPSETDG 1126
Qy 1133 YVAPLTCSPQPEYVNPQDVRRPQSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPA 1192
Db 1127 YVAPLTCSPQPEYVNPQDVRRPQSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVPA 1186
Qy 1193 FCGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPBERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPBERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVVPV 1261
Db 1247 EYLGLDVVPV 1255
RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antiseize;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX

PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowaert LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antiseize oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antiseize compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antiseize compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 97.9%; Score 6684; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;
Qy 1 MELAALCRWGLLLALLPFGAASQVCTGDMKRLPASPETHLDMLRHLVQGCQVVGQNL 60
Db 1 MELAALCRWGLLLALLPFGAASQVCTGDMKRLPASPETHLDMLRHLVQGCQVVGQNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIANQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQORNQVQDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGVLIQORNQVQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGSESDCCQSLTRTYCAGGCARCKGPLPENNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGSESDCCQSLTRTYCAGGCARCKGPLPENNFTVSF 240
Qy 241 WURPVKSASHLE-----KHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPE 292
Db 233 -----TDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPE 286
Qy 293 GRITFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLG 352
Db 287 GRITFGASCVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLG 346
Qy 353 MEHLREVRVTSANIQEFAGCKKIFGSLAFLESFSDGDPASNTAPLQPEQLQVFTLEEI 412
Db 347 MEHLREVRVTSANIQEFAGCKKIFGSLAFLESFSDGDPASNTAPLQPEQLQVFTLEEI 406
Qy 413 TGYLYISAWPDSPLDLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRGLSGLA 472
Db 407 TGYLYISAWPDSPLDLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRGLSGLA 466
Qy 473 LIHNTHLCFVHTVPMDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 532
Db 467 LIHNTHLCFVHTVPMDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLQGLPREYNARHCLPCHPECOFQNGSVTCFGEADQOVA 592

Db	527	QCNCVQFLRGQCEVECRVLOGLPREYNARHCLPCHPECOQONGSVTCFGEADQV	586
Qy	593	CAHYKDPFCVACPSGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAQ	652
Db	587	CAHYKDPFCVACPSGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAQ	646
Qy	653	RASPLTSIVSAVVGILLVVLGVVGGILIKRQOKIRKYTMRLLOETELVEPLTPSGAM	712
Db	647	RASPLTSIISAVVGILLVVLGVVGGILIKRQOKIRKYTMRLLOETELVEPLTPSGAM	706
Qy	713	PNQAOIRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE	772
Db	707	PNQAOIRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKE	766
Qy	773	ILDEAYNAGVGSPPVSRLLGLICTSTVOLVTLQMLPYGCLLDHVRENRRGLSQDLLNWC	832
Db	767	ILDEAYNAGVGSPPVSRLLGLICTSTVOLVTLQMLPYGCLLDHVRENRRGLSQDLLNWC	826
Qy	833	MOIAGMSYLEVDRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP	892
Db	827	MOIAGMSYLEVDRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP	886
Qy	893	KWMALESILRRFFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI	952
Db	887	KWMALESILRRFFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI	946
Qy	953	CTIDVYMIWVKWIMIDSECRPRFRELVSFSEFMRDQRFVVIQNEDLGPASPLDSTFYR	1012
Db	947	CTIDVYMIWVKWIMIDSECRPRFRELVSFSEFMRDQRFVVIQNEDLGPASPLDSTFYR	1006
Qy	1013	SLLEDMDGLVDAREYLVPOGFCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS	1072
Db	1007	SLLEDMDGLVDAREYLVPOGFCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPS	1066
Qy	1073	EEAPRSLAPSEAGSDVDFDGLQMGAAKGLQSLPTHDPSPLOYSDDPTVPLPSETDG	1132
Db	1067	EEAPRSLAPSEAGSDVDFDGLQMGAAKGLQSLPTHDPSPLOYSDDPTVPLPSETDG	1126
Qy	1133	YVAPLTCSPQPEYNQDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVA	1192
Db	1127	YVAPLTCSPQPEYNQDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVA	1186
Qy	1193	FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTAE	1252
Db	1187	FGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTAE	1246
Qy	1253	EYLGLDVVPV 1261	
Db	1247	EYLGLDVVPV 1255	
RESULT 12			
AAE20479			
ID	AAE20479	standard; Protein; 1255 AA.	
AC	AAE20479;		
DT	01-JUL-2002	(first entry)	
DE	Human Her-2/neu protein.		
KW	Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;		
OS	human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.		
FT	Homo sapiens.		
FT	Key	Location/Qualifiers	
FT	Region	1021..1030	
FT		/note= "Naturally processed HLA-B44-restricted epitope"	
PN	WO200214503-A2.		
XX			

PD	21-FEB-2002.		
XX	14-AUG-2001; 2001WO-US41733.		
PF	14-AUG-2000; 2000US-225152P.		
PR	28-SEP-2000; 2000US-236428P.		
PR	21-FEB-2001; 2001US-270520P.		
XX	(CORI-) CORIXA CORP.		
PA	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;		
PI	McNeill PD, Vedwick TS;		
XX	WPI; 2002-280758/32.		
DR	N-PSDB; AAD32743.		
XX	Novel isolated Her-2/Neu polypeptide composition useful for therapy,		
PT	prevention and diagnosis of cancer, preferably breast cancer		
XX	Disclosure; Page 114-117; 129pp; English.		
PS	The invention relates to an isolated Her-2/Neu polypeptide composition		
CC	effective for eliciting an immune response. The invention is useful for		
CC	eliciting an immune response in a patient, where the patient is human		
CC	leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.		
CC	The composition is useful for the therapy and diagnosis of cancer,		
CC	preferably breast cancer, in pharmaceutical compositions, e.g., vaccine		
CC	and other compositions for the diagnosis, prevention and treatment of		
CC	human malignancies, for stimulating and/or expanding T cells specific for		
CC	Her-2/Neu polypeptide and for inhibiting the development of cancer in a		
CC	patient. The invention is useful for stimulating a T cell response in a		
CC	human patient, as probe or primer for nucleic acid hybridisation, to		
CC	selectively form duplex molecules with complementary stretches of the		
CC	entire Her-2/Neu gene or gene fragments of interest, to isolate a full		
CC	length gene from a suitable library, and to direct expression of a		
CC	polypeptide in appropriate host cells. The composition is useful in		
CC	prophylactic or therapeutic applications and for the treatment of cancer,		
CC	preferably for the immunotherapy of breast cancer and other Her-2/Neu-		
CC	associated malignancies. The invention is useful in gene therapy. The		
CC	present sequence is human Her-2/neu protein.		
XX	Sequence 1255 AA;		
SQ	Query Match 97.9%; Score 6684; DB 23; Length 1255;		
	Best Local Similarity 97.6%; Pred. NO. 0;		
	Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;		
Qy	1	MELAAALCRWGLLLALLPPGAASCTCTGDMKRLPASPETHLDMRLHYQGVVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASCTCTGDMKRLPASPETHLDMRLHYQGVVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG	120
Qy	121	DPLNNTTPTVTGASPGGLREQLRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA	180
Db	121	DPLNNTTPTVTGASPGGLREQLRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA	180
Qy	181	LTLIDTNRSRACHPCSPMKCGRCWSESSEDCQSLTRTVCCAGGCARCKGPLPFNNVTSF	240
Db	181	LTLIDTNRSRACHPCSPMKCGRCWSESSEDCQSLTRTVCCAGGCARCKGPLPFNNVTSF	240
Qy	241	WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDFTFESMPNPE	292
Db	233	-----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFTFESMPNPE	286
Qy	293	GRYTFGASCVTACPYNYLSTDVGSCITLVCPHLNQEVTAEADGTQRCCKSKPCARVCYGLG	352
Db	287	GRYTFGASCVTACPYNYLSTDVGSCITLVCPHLNQEVTAEADGTQRCCKSKPCARVCYGLG	346
Qy	353	MEHLREVRVAVTSANIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLOVFTLBEI	412

Db 347 MEHLREAVTSANIQEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEQLQVFETLEBI 406
Qy 413 TGVLYISAWPDSLPDLSVFQNLQVIRGRILHNGAVSLTLQGLGISWGLRSLRELGSGLA 472
Db 407 TGVLYISAWPDSLPDLSVFQNLQVIRGRILHNGAVSLTLQGLGISWGLRSLRELGSGLA 466
Qy 473 LIHNTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHMGPGPT 532
Db 467 LIHNTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGACHOLCARGHMGPGPT 526
Qy 533 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCYA 592
Db 527 QCVNCSQFLRGQECVEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCYA 586
Qy 593 CAHYKDPFCVACRPSGVKPDLSYMPIWKFPEEGACQPCINCHTSCVDLDDKGCPEAQ 652
Db 587 CAHYKDPFCVACRPSGVKPDLSYMPIWKFPEEGACQPCINCHTSCVDLDDKGCPEAQ 646
Qy 653 RASPLTSIVSAVVGILLVVLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 712
Db 647 RASPLTSIIISAVVGILLVVLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAM 706
Qy 713 PNOAQRILKETELRKVKVLGSGAFCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQRILKETELRKVKVLGSGAFCTVYKGIWIPGENVKIPVAIKVIRENTSPKANKE 766
Qy 773 ILDEAYVMAGVSPYVRLIGLICLTSTVOLVTOLMPYGCCLDHVRENRLGSGODLLNWC 832
Db 767 ILDEAYVMAGVSPYVRLIGLICLTSTVOLVTOLMPYGCCLDHVRENRLGSGODLLNWC 826
Qy 833 MQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 892
Db 827 MQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 886
Qy 893 KWALESILRRRFTHOSDVMSYGVTVWELMTGCAKPYDGIIPAREIPDLLEKGERLPQPP 952
Db 887 KWALESILRRRFTHOSDVMSYGVTVWELMTGCAKPYDGIIPAREIPDLLEKGERLPQPP 946
Qy 953 CTIDVYMIWVKCMIDSECRPRELVSFSESRMARDPQRFVVIQNEGLGASPLDSTFYR 1012
Db 947 CTIDVYMIWVKCMIDSECRPRELVSFSESRMARDPQRFVVIQNEGLGASPLDSTFYR 1006
Qy 1013 SLEDDDDMGDLVDAEYLYVPOQFFCPDPAPGAGVHHRSSSTRSGGDLTLGLEPS 1072
Db 1007 SLEDDDDMGDLVDAEYLYVPOQFFCPDPAPGAGVHHRSSSTRSGGDLTLGLEPS 1066
Qy 1073 EEEAPSPAPSEGAGSDVFDGLGMAAGLQSLPHTDPSLQRYSEDTVPLPSETDG 1132
Db 1067 EEEAPSPAPSEGAGSDVFDGLGMAAGLQSLPHTDPSLQRYSEDTVPLPSETDG 1126
Qy 1133 YVAPLTCSQPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFA 1192
Db 1127 YVAPLTCSQPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFA 1186
Qy 1193 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFNLYWDQDPPERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTPOGGAAPQPPHPPAFSPAFNLYWDQDPPERGAPPSTFKGTPTAENP 1246
Qy 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255

RESULT 13

AA511143

ID AA511143 standard; Protein; 1255 AA.

XX AC AA511143;

XX 17-JUN-2002 (first entry)

DT Human Her-2/neu oncogene-encoded p185 glycoprotein.

DE

XX

XX

XX

XX

XX

XX

XX

XX

Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
tyrosine kinase; receptor; c-erbB2; gene therapy.

Homo sapiens.

Key Location/Qualifiers
Domain 1..653
/note= "extracellular domain"
Domain 676..1255
/note= "intracellular domain"
Domain 990..1255
/note= "phosphorylation domain"

W0200212341-A2.

PD 14-FEB-2002.

PF 03-AUG-2001; 2001WO-US24283.

PR 03-AUG-2000; 2000US-0632507.

PA (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Cheever MA, Gheysen D;

DR WPI; 2002-241743/29.

DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
or enhancing an immune response to the protein, has Her-2/neu
extracellular domain fused to Her-2/neu intracellular or
phosphorylation domain

PS Claim 68; Fig 7; 14pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
or c-erbB2), an oncogenic self-protein and target for anti-cancer
vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
in a variety of cancers, including breast, ovarian, colon, lung and
prostate cancer. Her-2/neu is a member of the tyrosine kinase
family of receptor-like glycoproteins. It comprises an extracellular
domain with homology to the epidermal growth factor receptor
(EGFR), a highly hydrophobic transmembrane domain and a C-terminal
intracellular domain that also shows homology to EGFR. Its
overexpression correlates with a poor prognosis in breast and
ovarian cancers. The invention provides Her-2/neu fusion
proteins, nucleic acids encoding them, viral vectors, and vaccines
comprising the fusion proteins or nucleic acid molecules. In
preferred fusion proteins, the extracellular domain of a Her-2/neu
protein is fused to a Her-2/neu intracellular domain or
phosphorylation domain (or its DeltaAPP fragment). An immune
response to Her-2/neu protein is elicited or enhanced by
administering the fusion protein in the form of a vaccine, or by
transfecting cells of an animal ex vivo with a nucleic acid
encoding the fusion protein, and delivering the transfected cells
to the animal. The fusion proteins, nucleic acids, and isolated
specific T-cells are useful for inhibiting the development of a
cancer, especially breast, ovarian, colon, lung or prostate cancer
in a patient. T cells that specifically react with a Her-2/neu
fusion protein can be used to remove tumour cells from a sample in
order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

Qy 1 MELAALCRWGLLLALLPGCAASTQVCTGTDMLRLPASPTHLDMLRHLYQCVVQGNL 60

Db 1 MELAALCRWGLLLALLPGCAASTQVCTGTDMLRLPASPTHLDMLRHLYQCVVQGNL 60

```
QY 61 ELTYLPTNASLSFLODIEQVGYVLI AHNVQVFLORLRIYVRGTQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLSFLODIEQVGYVLI AHNVQVFLORLRIYVRGTQLFEDNYALAVLNG 120
QY 121 DPLNNTTPVTGASPGGLREQLRSLEILKGGVLQRPOLCYQDTILWKDIFHKQNQLA 180
DB 121 DPLNNTTPVTGASPGGLREQLRSLEILKGGVLQRPOLCYQDTILWKDIFHKQNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKGSRGWGESSEDQSLTRTVACGACRCKGPLE----- 232
DB 181 LTLIDTNRSRACHPCSPCKGSRGWGESSEDQSLTRTVACGACRCKGPLE----- 232
QY 241 WLRVPKVSASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292
DB 241 WLRVPKVSASHLE-----KHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPE 292
QY 293 GRYTEGASCVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
DB 293 GRYTEGASCVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLG 352
QY 353 MEHLREVRRAVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQBPQLQVFTLEBI 412
DB 353 MEHLREVRRAVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQBPQLQVFTLEBI 412
QY 413 TGYLYISAWPDSLPDLSVFNQNLQVIRGRILHNGAYSLTLOGLIGLWGLRLSRLGSLA 472
DB 413 TGYLYISAWPDSLPDLSVFNQNLQVIRGRILHNGAYSLTLOGLIGLWGLRLSRLGSLA 472
QY 473 LIHNTHLCLFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPGPT 532
DB 473 LIHNTHLCLFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPGPT 532
QY 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQVCA 592
DB 533 QCVNCSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQVCA 592
QY 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCCPAEQ 652
DB 593 CAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKCCPAEQ 652
QY 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRRQOKIRKYTWRLLOQETELVEPLTPSGAM 712
DB 653 RASPLTSIVSAVVGILLVVLGVVFGILIKRRQOKIRKYTWRLLOQETELVEPLTPSGAM 712
QY 713 PNOAQMRLKETELRKVKVLSGARGTYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
DB 713 PNOAQMRLKETELRKVKVLSGARGTYKGIWIPDGENVKIPVAIKVLRNTPSKANKE 772
QY 773 ILDEAYVMAGVSPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRRGLSGODLLNWC 832
DB 773 ILDEAYVMAGVSPVSRLLGICLTSTVOLVTQMLPYGCLLDHVRNRRGLSGODLLNWC 832
QY 833 MOIAGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPI 892
DB 833 MOIAGMSYLEDVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPI 892
QY 893 KMALESILRRFRTHQSDVNSVGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
DB 893 KMALESILRRFRTHQSDVNSVGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
QY 953 CTIDVYMIWKCMIIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYR 1012
DB 953 CTIDVYMIWKCMIIDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGASPLDSTFYR 1012
QY 1013 SLLEDDDMGDLVDAEYLVPOQGFPCPDPAAGAGMWHRRSSRSGGDLTLGLBPS 1072
DB 1013 SLLEDDDMGDLVDAEYLVPOQGFPCPDPAAGAGMWHRRSSRSGGDLTLGLBPS 1072
QY 1073 EEEAPRSPAPSEAGSDVFDGDLGMGAAGLQSLPHTHDPSPLOQRYSEDPTVPLPSETDG 1132
DB 1073 EEEAPRSPAPSEAGSDVFDGDLGMGAAGLQSLPHTHDPSPLOQRYSEDPTVPLPSETDG 1132
QY 1133 YVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKVFA 1192
```

```
DB 1127 YVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERPKTILSPGKNGVVKVFA 1186
QY 1193 FCGAVENPEYLTPOCGAARPOPHPPPAFPAFONLYYWDODPPERGAPPSTFKGTPTAENP 1252
DB 1187 FCGAVENPEYLTPOCGAARPOPHPPPAFPAFONLYYWDODPPERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
DB 1247 EYLGLDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; Chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI; 2002-280741/32.
XX
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 97.9%; Score 6684; DB 23; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1239; Conservative 2; Mismatches 6; Indels 22; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASVQCTGTDMLRLPASPTHDLMLRHLHYGQCVVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASVQCTGTDMLRLPASPTHDLMLRHLHYGQCVVQGNL 60
QY 61 ELTYLPTNASLSFLODIEQVGYVLI AHNVQVFLORLRIYVRGTQLFEDNYALAVLNG 120
```

Db 61 ELTYLPTNASLSFLQDIOQVGVLIHNVQVPLQRLIRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGSESSDCOSLTRTVCAGGCARCKGPLPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMCKGRCWGSESSDCOSLTRTVCAGGCARCKGPLP----- 232
QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 292
Db 233 -----TDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE 286
QY 293 GRYTTCASCVTACPNYLSLTDVGSCTLVCPLNHOEVTAEQGTORCEKSKPCARVCYGLG 352
Db 287 GRYTTCASCVTACPNYLSLTDVGSCTLVCPLNHOEVTAEQGTORCEKSKPCARVCYGLG 346
QY 353 MEHLREVAVTSANIOEFAGCKKI FGS LAF LPE SFDGDPASNTAPLOPEQLQVFELEEI 412
Db 347 MEHLREVAVTSANIOEFAGCKKI FGS LAF LPE SFDGDPASNTAPLOPEQLQVFELEEI 406
QY 413 TGYLYISAWPDSLPLDSVFNQLQVIRGRILHNGAYS LTLQGLGISWGLRSLRELGLA 472
Db 407 TGYLYISAWPDSLPLDSVFNQLQVIRGRILHNGAYS LTLQGLGISWGLRSLRELGLA 466
QY 473 LIHNTHLFCVHTVPMQDLFRNPHQALHTANRPEDECYGEGLACHOLCARGHCWPGPT 532
Db 467 LIHNTHLFCVHTVPMQDLFRNPHQALHTANRPEDECYGEGLACHOLCARGHCWPGPT 526
QY 533 QCVNCSQFLRGQECVECEVRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 592
Db 527 QCVNCSQFLRGQECVECEVRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCV 586
QY 593 CAHYKDPFCVACRPSGVKPDLSYMPIWKFPEDEGACQPCNCTHSCVDLDDKGGPABQ 652
Db 587 CAHYKDPFCVACRPSGVKPDLSYMPIWKFPEDEGACQPCNCTHSCVDLDDKGGPABQ 646
QY 653 RASPLTSIVS VVGI LLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAM 712
Db 647 RASPLTSIIISAVVGI LLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAM 706
QY 713 PNOAQMRLIKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 772
Db 707 PNOAQMRLIKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKE 766
QY 773 ILDEAYVMAGVSPYVSRLLGICLTSTVQLVTOLMPYGCGLLHVRENRRGLSGQDLLNWC 832
Db 767 ILDEAYVMAGVSPYVSRLLGICLTSTVQLVTOLMPYGCGLLHVRENRRGLSGQDLLNWC 826
QY 833 MOIAGMSYLEVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVP 892
Db 827 MOIAGMSYLEVRLVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVP 886
QY 893 KMALESILRRRFTHOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 887 KMALESILRRRFTHOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPI 946
QY 953 CTIDVYIMVWKMIIDSECRPRFRELVSFERNARDPQRFVJQNEEDLGPASPLDSTFYR 1012
Db 947 CTIDVYIMVWKMIIDSECRPRFRELVSFERNARDPQRFVJQNEEDLGPASPLDSTFYR 1006
QY 1013 SULLEDDMDGLVDABEYLVPOQGFCCPDAPAGAGVHHRSSSTRSGGDLTLGLEPS 1072
Db 1007 SULLEDDMDGLVDABEYLVPOQGFCCPDAPAGAGVHHRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDG 1132
Db 1067 EEEAPRSLAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDTVPLPSETDG 1126
QY 1133 YVAPLTCSQPEYVNVQDVRPQPPSPREGPLPAARPAGATLERPKT LSPGKNGVWKDVFA 1186

Db 1127 YVAPLTCSQPEYVNVQDVRPQPPSPREGPLPAARPAGATLERPKT LSPGKNGVWKDVFA 1186
QY 1193 FCGAVENPEYLTPOCGAARPOPHPPAFSFAFNLYYMDODPBERGAPPSTFKGTPTAENP 1252
Db 1187 FCGAVENPEYLTPOCGAARPOPHPPAFSFAFNLYYMDODPBERGAPPSTFKGTPTAENP 1246
QY 1253 EYLGLDVPV 1261
Db 1247 EYLGLDVPV 1255
RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
AC AAR39568;
XX XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93MO-USO1055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
DR WPI: 1993-272889/34.
DR N-PSDB; AAQ46083.
XX
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
Query Match 97.3%; Score 6641; DB 14; Length 1433;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1232; Conservative 4; Mismatches 11; Indels 22; Gaps 2;
QY 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLPASPETHLDMLRHLYQGCQVVOGNI 60
Db 1 MELAALCRWGLLALLPPGAASSTQVCTGDMKRLPASPETHLDMLRHLYQGCQVVOGNI 60
QY 61 ELTYLPTNASLSFLQDIOQVGVLIHNVQVPLQRLIRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOQVGVLIHNVQVPLQRLIRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGSLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGSESSDCOSLTRTVCAGGCARCKGPLPFNNFTVSF 240
Db 181 LTLIDTNRSRACHPCSPMCKGRCWGSESSDCOSLTRTVCAGGCARCKGPLP----- 232

QY 241 WLRVPKVSASHLE-----KHSDCLACLFHNSHGI CELHCPALVTVNTDTFESMPNPE 292
Db :
233 -----TDCCHQCAAGCTGPKHSDCLACLFHNSHGI CELHCPALVTVNTDTFESMPNPE 286
QY 293 GRYTFGASCVTACPNYLSTVDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLG 352
Db :
287 GRYTFGASCVTACPNYLSTVDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLG 346
QY 353 MEHLREVRVAVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELQOVFETLEBI 412
Db :
347 MEHLREVRVAVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELQOVFETLEBI 406
QY 413 TGLYLISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLTLOGLGISWLGILRSRLRELSGLA 472
Db :
407 TGLYLISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLTLOGLGISWLGILRSRLRELSGLA 466
QY 473 LIHNNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPPT 532
Db :
467 LIHNNTHLSFVHTVPWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPPT 526
QY 533 QCVNCSQFLRGOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCV 592
Db :
527 QCVNCSQFLRGOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCV 586
QY 593 CAHYKDPPEFCVARGCPGKPDLSYMPINWKFDEEGACQPCPINCTHSCVDLDDKGCPEAQ 652
Db :
587 CAHYKDPPEFCVARGCPGKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQ 646
QY 653 RASPLTSTIVSAVVGILLVVVLGVWFGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAM 712
Db :
647 RASPLTSTIVSAVVGILLVVVLGVWFGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAM 706
QY 713 PNOQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPEKANKE 772
Db :
707 PNOQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPEKANKE 766
QY 773 ILDEAYVMAGVSPVSVRLGLCLTSTVOLVTQLMPYGCCLLDHVRENRRGLSGQDLLNWC 832
Db :
767 ILDEAYVMAGVSPVSVRLGLCLTSTVOLVTQLMPYGCCLLDHVRENRRGLSGQDLLNWC 826
QY 833 MOIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 892
Db :
827 MOIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 886
QY 893 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPI 952
Db :
887 KMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPI 946
QY 953 CTIDVYMIWVKCWMIDSECRPRFRELUSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db :
947 CTIDVYMIWVKCWMIDSECRPRFRELUSEFSRMDPQRFVVIQNEIDLGPASPLDSTFYR 1006
QY 1013 SLLEDDDMGDLVDAEYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPS 1072
Db :
1007 SLLEDDDMGDLVDAEYLVPOQGFCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPS 1066
QY 1073 EEEAPRSLAPSEAGSDVFDGDLGMAAKGLSLPHTDPSLQRYSEDPTVPLPSETDG 1132
Db :
1067 EEEAPRSLAPSEAGSDVFDGDLGMAAKGLSLPHTDPSLQRYSEDPTVPLPSETDG 1126
QY 1133 YVAPLTCSPQPEYVNDVVRQPPSPREGPLPAARPAGATLERAKTILSPGKGVVVDVFA 1192
Db :
1127 YVAPLTCSPQPEYVNDVVRQPPSPREGPLPAARPAGATLERAKTILSPGKGVVVDVFA 1186
QY 1193 FGGAVENPEYLTPOGGAAPQHPAPSPAFDNLVYMDODPPERCAPSTFKGTPTAENP 1252
Db :
1187 FGGAVENPEYLTPOGGAAPQHPAPSPAFDNLVYMDODPPERCAPSTFKGTPTAENP 1246
QY 1253 EYLGLDVVP 1261
Db :
1247 EYLGLDVVP 1255

Search completed: July 22, 2003, 08:40:55
Job time : 43.1589 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-250-264-14
Perfect score: 6839
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYLCGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6702	98.0	1255	1 A24571	protein-tyrosine k
2	5895	86.2	1260	1 TVRTNU	protein-tyrosine k
3	5880.5	86.0	1254	2 I48161	p-185 precursor
4	3116	45.6	1210	1 GQHUE	epidermal growth f
5	3092	45.2	1210	2 A53183	epidermal growth f
6	3071.5	44.9	1223	1 TVCHLV	epidermal growth f
7	2971.5	43.4	1308	2 A47253	epidermal growth f
8	2674	39.1	1166	1 S06142	protein-tyrosine k
9	2391.5	35.0	1342	2 A36223	kinase-related tra
10	2306.5	33.7	1339	2 JQ4387	epidermal growth f
11	1766.5	25.8	698	1 TVFVLV	protein-tyrosine k
12	1703	24.9	604	1 TVYUHV	protein-tyrosine k
13	1647	24.1	544	2 S35745	protein-tyrosine k
14	1640	24.0	545	2 S00727	protein-tyrosine k
15	1623	23.7	540	2 B44776	kinase-related tra
16	1621	23.7	540	1 TVFVEB	protein-tyrosine k
17	1600.5	23.4	1330	1 GQFFE	epidermal growth f
18	1484	21.7	644	2 A36325	epidermal growth f
19	1265	18.5	1323	2 S88257	protein let-23 (im
20	1265	18.5	1374	2 S70712	protein-tyrosine k
21	1171	17.1	1369	2 S70713	protein-tyrosine k
22	1146	16.8	1717	1 A45558	epidermal growth f
23	1103	16.1	527	2 A42032	epidermal growth f
24	945.5	13.8	843	2 A27131	epidermal growth f
25	866.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	733	10.7	1363	2 T32220	insulin-like growth
28	721.5	10.5	1372	2 A34157	insulin receptor p
29	719	10.5	1383	2 A36080	insulin receptor p

ALIGNMENTS

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein e
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.
Nature 319, 230-234, 1986
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; MID:g31197; PIDN:CAA27060.1; PID:g31198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; MID:g182163; PIDN:AAA35808.1; PID:g553282
R:Coussens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; MID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; MID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:L29395; MID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

30 718 10.5 1382 1 INHUR insulin receptor p
31 691 10.1 1607 2 T43212 insulin-like growth
32 687 10.0 1477 2 T18534 protein-tyrosine k
33 685.5 10.0 1300 2 A36502 insulin receptor-r
34 675 9.9 1268 2 B36502 insulin receptor-r
35 649 9.5 1367 1 IGHUR1 insulin-like growth
36 635 9.3 1371 2 A33837 insulin-like growth
37 620 9.1 1390 2 T30346 insulin receptor -
38 611.5 8.9 2148 1 A56081 insulin receptor -
39 606 8.9 2101 2 S57245 insulin receptor (
40 594.5 8.7 987 2 A54092 protein-tyrosine k
41 589.5 8.6 977 2 S49004 tyrosine kinase Mp
42 588 8.6 1114 1 S05582 protein-tyrosine k
43 586 8.6 1091 2 S33596 protein-tyrosine k
44 584.5 8.5 976 2 A36355 protein-tyrosine k
45 583.5 8.5 952 2 I50612 protein-tyrosine k

A:Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C:Genetics:
A:Gene: GDB:ERBB2; NGL; NEU; HER-2
A:Cross-references: GDB:120613; OMIM:164870
A:Map position: 17q21.1-17q21.1
A:Introns: 25/1; 75/3; 147/1; 883/3
A>Note: the list of introns is incomplete
C:Function:
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F:22-653/Domain: extracellular #status predicted <EXT>
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F:654-675/Domain: transmembrane #status predicted <TM>
F:676-1255/Domain: intracellular #status predicted <INT>
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
F:68,124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:753/Active site: Lys #status predicted
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6702; DB 1; Length 1255;
Best Local Similarity 98.3%; Pred. No. 8.3e-268; Indels 6; Gaps 1;
Matches 1240; Conservative 3; Mismatches 12;

QY	1	MELAAALCRNGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL	60
DB	1	MELAAALCRNGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL	60
QY	61	ELTYLPTNASLSFLQDIEQVGVLLIAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG	120
DB	61	ELTYLPTNASLSFLQDIEQVGVLLIAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG	120
QY	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQVQDITLWKDIFPHKNQOLA	180
DB	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQVQDITLWKDIFPHKNQOLA	180
QY	181	LTLIDNTRACHPCSPKMGSCWGESSEDCOSLTRTVCAGGCARCKGFLPTDCCHEQC	240
DB	181	LTLIDNTRACHPCSPKMGSCWGESSEDCOSLTRTVCAGGCARCKGFLPTDCCHEQC	240
QY	241	AAGCTGPKHSDCLACILFHNSGICELHCPALVFNNTVFWLRVVKVASHLRYTFGAS	300
DB	241	AAGCTGPKHSDCLACILFHNSGICELHCPALVFNNTVFWLRVVKVASHLRYTFGAS	294
QY	301	CVTACPNYVLTSDVGSCTVLCPLHNOBVTAEQDQTCCKSKPCARVCYGLGMEHLREVR	360
DB	295	CVTACPNYVLTSDVGSCTVLCPLHNOBVTAEQDQTCCKSKPCARVCYGLGMEHLREVR	354
QY	361	AVTSANIQEFAGCKKIFGSLAFUPESFDGDPASNTAPLOEQVQVETLEEITGYLYISA	420
DB	355	AVTSANIQEFAGCKKIFGSLAFUPESFDGDPASNTAPLOEQVQVETLEEITGYLYISA	414
QY	421	WDSLPLDSVFQNLQVIRGRLHNGAYSILTLQGLISWGLRLSLRELGLALIHNTHL	480
DB	415	WDSLPLDSVFQNLQVIRGRLHNGAYSILTLQGLISWGLRLSLRELGLALIHNTHL	474
QY	481	CFVHTVPDQLFRNPQALLHTANRPDECVGSLACHOLCARHGCWGPCPTQVCNCSOF	540
DB	475	CFVHTVPDQLFRNPQALLHTANRPDECVGSLACHOLCARHGCWGPCPTQVCNCSOF	534
QY	541	LRGQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDP	600
DB	535	LRGQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPGPADQCACAHYKDP	594
QY	601	FCVARCPGKPDLSYMPITWKPDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSI	660

DB	595	FCVARCPGKPDLSYMPITWKPDEEGACQPCPINCTHSCVDLDDKGCFAEORASPLTSI	654
QY	661	VSNAVGIILLVVLGVVFGILLIKRQOKIRKYTMRRLLQETELVEPLTPSCAMPNOQMRI	720
DB	655	ISAVGILLVVLGVVFGILLIKRQOKIRKYTMRRLLQETELVEPLTPSCAMPNOQMRI	714
QY	721	LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM	780
DB	715	LKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM	774
QY	781	AGVGSYVSRLLIGLICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLLNWCWOIAKMS	840
DB	775	AGVGSYVSRLLIGLICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGQDLLNWCWOIAKMS	834
QY	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDIDETEHADGGKVPICKMALES	900
DB	835	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDIDETEHADGGKVPICKMALES	894
QY	901	LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDYYMI	960
DB	895	LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDYYMI	954
QY	961	MVKCMIDSECRPRFRELVSERFARMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDDM	1020
DB	955	MVKCMIDSECRPRFRELVSERFARMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDDM	1014
QY	1021	GDLDVAEEYLVPOQGFCCPDPAAGAGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP	1080
DB	1015	GDLDVAEEYLVPOQGFCCPDPAAGAGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP	1074
QY	1081	LAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTC	1140
DB	1075	LAPSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTC	1134
QY	1141	POPEYVNPQDVPRPQPPSPREGPLPAARPAAGATLERAKTLSPGKQGVVQVAFGGA	1200
DB	1135	POPEYVNPQDVPRPQPPSPREGPLPAARPAAGATLERAKTLSPGKQGVVQVAFGGA	1194
QY	1201	EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTAENPEYLGUDVP	1260
DB	1195	EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTAENPEYLGUDVP	1254
QY	1261	V 1261	
DB	1255	V 1255	

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a
2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <NAS>
A>Note: authors translated the codon GCA for residue 25 as Val
A:Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;731.191.263.535.576.634/Binding site: carboxylate (Asn) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.2%; Score 5895; DB 1; Length 1260;
Best Local Similarity 86.4%; Pred. No. 1e-234;
Matches 1091; Conservative 52; Mismatches 112; Indels 8; Gaps 3;

Qy 1 MELAALCRWGLLLALLPFGAASSTVCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60
Db 4 MELAALCRWGLLLALLPFGAAGTCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 63

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVGTQLFEDNYVALAVLNG 120
Db 64 ELTYVPANASLSFLQDIOEVQGYMLIAHQVQVPLQRLIRVGTQLFEDKYALAVLNR 123

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHKNNQL 179
Db 124 DPQNVAASTFGRTPEGLRELQRLSLTEILKGGVLIQORNPOLCYODMVLWKDVFHKNNQL 183

Qy 180 ALTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVGAGGCARCKGRLPTDCCHEQ 239
Db 184 APVDIDTNRSRACHPCSPACKDNHCWGESPEDCQILTGTCTGCGCARCKGRLPTDCCHEQ 243

Qy 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVFLWVQVKSASHLEYRTFGA 299
Db 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNP-----GRYTFGA 297

Qy 300 SCVTACPNYLSLTDVSGCTLCPLHQBVTAEADGTQRCCKSKPCARCYGLGHEHREV 359
Db 298 SCVTCPNYLSLTVSGCTLCVPPNQVETAEADGTQRCCKSKPCARCYGLGHEHRLGA 357

Qy 360 RAVTSANIQEPAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFTLEITGYLYIS 419
Db 358 RAITSNDVQEPDGCKKIFGSLAFIPESFDGDPSSGIAPLRPEQLQVFTLEITGYLYIS 417

Qy 420 AWPDSLPLSVFQNLQVIRGILHNGAYSLTQLGTSWGLSRLSRELGSGLAIHNNHT 479
Db 418 AWPDSLRLSVFQNLRIIRGILHDGAYSLTQLGTHSLGSLRSLRELGSGLAIHNNH 477

Qy 480 LCFVHTVPWDLFRNPHQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPPTQCVNCS 538
Db 478 LCFVHTVPWDLFRNPHQALLHSGNRPEEDLCVSSGLVCNLSLCAHGCWGPPTQCVNCS 537

Qy 539 QFLRGQCEVERVLQGLPREYVNHARHCLPCHPECQPNQSVTCFGEADQCACAHYKD 598
Db 538 HFLRGQCEVERVWGLPREYVSDKRLCPCHPECQPNQSVTCFGEADQCACAHYKD 597

Qy 599 PPECVAPCSGVKPDLSYMPYKPFDEGACQPCPINCSTHSCVDLDKQCPAQRASPLT 658
Db 598 SSSVAPCSGVKPDLSYMPYKPYDEEGICQPCPINCSTHSCVDLDKQCPAQRASPT 657

Qy 659 SIYSAVGILLVVLGVVFGILIKRQOKIRKYTMRLQQLQTELVEPLTPSGAMPNQAQM 718
Db 658 FIATVEGVLFLVLVVGILLIKRQOKIRKYTMRLQQLQTELVEPLTPSGAMPNQAQM 717

Qy 719 RILKETELRKVKVLGSGAGFTYKGIWIPQENVKIPVAIKVLRNTPSPKANKEILDEAY 778
Db 718 RILKETELRKVKVLGSGAGFTYKGIWIPQENVKIPVAIKVLRNTPSPKANKEILDEAY 777

Qy 779 VMAGVGSPPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNHRGLGSQDLLNWCQMIAKG 838
Db 778 VMAGVGSPPYVSRLLIGICLTSTVQLVTOIMPYGCLLDHVRNHRGLGSQDLLNWCQVIAKG 837

Qy 839 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 898

Db 838 MSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMAL 897

Qy 899 SILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVY 958

Db 898 SILRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVY 957

Qy 959 MIMVKCMWIDSECRPRFRELVSFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1018

Db 958 MIMVKCMWIDSECRPRFRELVSFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1017

Qy 1019 DMGDLVDAEYLVPOOGFFCDDPAPACAGMWHRRSSSTRSGGDLTLGLPSESEAPR 1078

Db 1018 DMGDLVDAEYLVPOOGFFCDDPAPACAGMWHRRSSSTRSGGDLTLGLPSESEAPR 1077

Qy 1079 SPLAPSEGAGSDVFDGDLGMAAGLQSLPLTHDPSPLORYSDPTVPLPSETDGYVAPLT 1138

Db 1078 SPLAPSEGAGSDVFDGDLGMAAGLQSLPLTHDPSPLORYSDPTVPLPSETDGYVAPLA 1137

Qy 1139 CSPQPEYVQPDVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAGGAVE 1198

Db 1138 CSPQPEYVQSEVQPPPLTPGGLPPVRPAGATLERPKTLSPGKNGVVKDVFAGGAVE 1197

Qy 1199 NPEYLTPOGGAAPQHPHPPAFSPAFDNLYYNDQDPPERGAPPSTFKGTPTAENPEYLGLD 1258

Db 1198 NPEYLVPREGTASPPHPSPAFPAFNDLYYNDQNSSEQPPSPNFEGTPTAENPEYLGLD 1257

Qy 1259 VPV 1261

Db 1258 VPV 1260

RESULT 3
148161
P-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: 148161
Gene 140, 251-255, 1994
R:Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Iehika
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: 148161; MUID:94193007; PMID:7908275
A:Accession: 148161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.0%; Score 5880.5; DB 2; Length 1254;
Best Local Similarity 86.1%; Pred. No. 4e-234;
Matches 1086; Conservative 60; Mismatches 108; Indels 7; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASSTVCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60
Db 1 MELAALCRWGLLLALLPFGASGTQVCTGTDMLRLPASPTHLDMLRLHYQCCVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVGTQLFEDNYVALAVLNG 120
Db 61 ELTYLPANATLSFLQDIOEVQGYMLIAHQVQVPLQRLIRVGTQLFEDKYALAVLNR 120

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHKNNQL 180
Db 121 DPLDNVTATGRTPEGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDVFHKNNQL 180

Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVGAGGCARCKGRLPTDCCHEQ 240
Db 181 PVDIDTNRSRACHPCSPACKDNHCWGESPEDCQTLTGTIAPRAVPAARARLPTDCCHEQ 240

R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5203-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744,'X',746-747 <RUS>
R;Mroczkowska, B.; Morig, G.; Cohen, S.
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity
A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain x
A;Reference number: A3331; MUID:90003233; PMID:2790960
A;Contents: annotation; internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:120610; OMIM:131550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
F;1-24/Domain: signal sequence #status predicted <Sig>
F;25-120/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <BE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <BE2>
F;646-668/Domain: transmembrane #status predicted <TM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-725/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;745/Active site: Lys #status experimental

Query Match 45.6%; Score 3116; DB 1; Length 1210;
Best Local Similarity 49.2%; Pred. No. 9e-121;
Matches 625; Conservative 179; Mismatches 355; Indels 112; Gaps 22;

Qy 11 LLLALLPPGAA--STQVCTGDMKRLPASPETHLDMRLHYQCQVQVQGNLEITYLPTN 68
Db 14 LLAALCPASRALEBKVKVCOGTSNKLTLQGTFFDHFLSLQRMFNCEVVLGNLEITYYQVN 73
Qy 69 ASLSFLQDIOEVQGYVLIHNOVQVPLRLRIVRGTLQLEDNAYALVLDNGDPLNNTTP 128
Db 74 YDLSFLTKIQEVAGYVLIANTVERIPLENLQIIRGNMYENSVALAVLSNYD----- 126
Qy 129 VTGASPGGLRELQURSILTEILKGGVLIQRNPOLCYQDTILWKQIFHKNNQLALTLIDTNR 188
Db 127 ---ANKTGLKELPMRNLEILHGAVERSNPNALCNVESIQWRDIVSSDFLSNMSMDQNH 183
Qy 189 SRACHPCSPMKSGRCWGESSEDCQSLTRTVACGCA-RCKGPLPTDCHEQCQAAGCTGP 247
Db 184 LGSQCQKDPSCPNCSGWCAGEENCQKLTIKCAQCSCGRCKRSPSCCHNQCAAGCTGP 243
Qy 248 KHSDDLACILHFNHSGICELCPALVFNNTVSWFLRVPKVSASHLEYRTGCASCVTCAPY 307
Db 244 RESDCLVCRFBREBATKOTCPPLMLYNPTY-----QMDVNPBGKYSFGATCKVKCKPR 297
Qy 308 NYLSTDVGSCTLVCPILHNOEVTADGTCRCEKSKPCARVCYGLGMEHLREVRVAVTSANI 367
Db 298 NYVVTDHGSCVRACGADSYEM-EEDGVRKCKCEGPCRKVCNGIGIGEFKDSLSINATNI 356
Qy 368 QEPAGCKKIFGSLAFIPESDGPASNTAPLQEQLOVFTLBEITGLYLYISAWPDSLPD 427
Db 357 KHFNKNTSISGDLHILPFAVRGDSFTHTPPLDPQELDKLVKTEITGFLLIQWPNRTD 416
Qy 428 LSVFQNLQVIRGRILHNGAVSLTLQGLIGISMLGRSLRELGSGLALIHNTLHCFVHTVP 487

Db 417 LHAFENLEIRGRTRKQHQGFSLAVVLSNITSLGRSLRSLKEISDGDVLIISGNKNLCYANTIN 476
Qy 488 WDOLFNRPHQALLHTANRPDECVGEGLACHQLCARGHCWGPPTQCVCNCSOFARGQECV 547
Db 477 WKKLFGTSGQTKIISNRGENSCATQVCHALCSPECCKWGPEDPCVSCRNVSREGREV 536
Qy 548 EECRVLOGLPREYINARHCLPCHBECOPQNSVTCFGEADOCVACAHYKDPFPFCVACRP 607
Db 537 DKCKLLEGEPRFVENSECICHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTPC 596
Qy 608 SGVKPDLSPYMPKWFPEDEGACQPCPNCTHSCVDLDDKGPQAEORASPLTSIVSAVVG- 566
Db 597 AGVNGENNTL-VWKYADAGHVCHLCHNCTYCTGCTGPGLEGCTNGPKTP--SIATGMVGA 653
Qy 667 --ILLVVVLGVVFGILIKRRQOKIKRYTMRRLLOETELVEPLTPSGAMPNOQMILKET 724
Db 654 LLLLVVALGIG--LFWRRHRIYVKTLRLRLQERELVEPLTPSGEAPNQALLILKET 710
Qy 725 ELRKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGV 784
Db 711 EFKIKIKVLGSGAFQTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMASVD 770
Qy 785 SPYVSRLLIGICTSTVOLVTOLMPYGLLDHVRNRRGLSGQDLLNWCNQTAKGMSYLED 844
Db 771 NPHVCRLLIGICTSTVOLITQLMPFGCLLDYVREHKONIGSQYLLNWCVCQIAKGMNYLED 830
Qy 845 VRLVHRDLAARNVLKSPNHWKITDFGLARLLDDIETEHADGGKVPKIMMALESILRRR 904
Db 831 RRLVHRDLAARNVLVKTPQHVKITDFGLAKULLGAEEKEYHAEGGKVPKIMMALESILHRI 890
Qy 905 FTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKC 964
Db 891 YTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSILEKGERLPQPPICITIDVYIMVKC 950
Qy 965 WMIDSECPREFRELVSFBSRMARDPQRFVIO-NEDLGASPLDSTFYRSLLEDDDDMDGL 1023
Db 951 WMIDADSRPFRELIIEFBSKARDPQRYLVIQGERMHLPSPDTSNFYRALMDEEDMDV 1010
Qy 1024 VDAEYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEAPEAPSLAP 1083
Db 1011 VDAEYLVPOQGF-----SSPSTSTPLLS 1036
Qy 1084 SEGAGSDVDFDGLGMGAAGLQSLTHDPSPQORYSEDPVPLPSET--DGYVAPLTCSP 1141
Db 1037 SLSATSN--NSTVACIDRNLQSCPIKEDSFQRYSSDPTGALTEDSIDDTFL-----P 1088
Qy 1142 QPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTSLSPKNGVVKDVFAGFGAVNPE 1201
Db 1089 VPEYINQ-SVPKRPAGSVQNPVYHNQPLNP-----APSRDPHYQD--PHSTAVGNPE 1137
Qy 1202 YL-TPOGGAAPQHPHPPAFSPAFDNLVYDQ-----DP-----PERGAPPSTFK 1244
Db 1138 YLNTVQ-----PTCVNSTFDSPAHWAAQKSHQISLDNPDYQDFFFPKEAPNGIFK 1189
Qy 1245 GTPTAENPEYL 1255
Db 1189 GS-TAENAEYL 1198

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C;Accession: A53183; A43816; S24942; A28941; S45325; I49643
R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
 A:Reference number: A43818; PMID:91232866; PMID:2030916
 A:Accession: A43818
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698
 R:Eisinger, D.P.; Sertero, G.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942
 A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <BIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; PMID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Patia, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse B
 A:Reference number: I49643; PMID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:g193001; PIDN:AAA53029.1; PID:g567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphopro
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:728-728/Domain: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45-2%; Score 3092; DB 2; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 8.7e-120;
 Matches 628; Conservative 171; Mismatches 363; Indels 116; Gaps 24;

Qy 11 LLLALLPGAA--STQVCTGDKMLRLPASBETHLDMLRLHYQGVQVQGNLELYLPTN 68
 Db 14 LTLTCAAGGAALEKKVCGQSNRLTQGTGFEDHFLSLQRMVNNCEVVLGNLEITYQVN 73
 Qy 69 ASLGFQIQIEVQGVYVLAHQVQVPLQRLIRVRGTQLFEDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLFLKTIQEVAGVIALNTVERIPLENQIIRGNALYENTYALAILSN----- 124
 Qy 129 VTGASPGGLRELQRLSRLTEILKGVLIQIRNPOLCYQDTILMKDI----FHKNNQALTLI 184
 Db 125 -YGNRTGLRELPMRLNQLLEILLIGAVRFNSNNPILCNMDTIQWRDIQVNVFMSNMSMDL--- 180
 Qy 185 PTNRSRACHPCSPMKSGRCHGESEDCSLTRTVCAGGCA-RCKGPLPTDCCHQCAAG 243
 Db 181 -QSHPSCFKCDPSPGNCGCGGGEENCKQTKIICAQOCSHRRCGRSPSCCHNQCAAG 239

Qy 244 CTGPKHSDCLACLHFNHSGICELHCPALVFNFTVSWLVRPKVSASHLRYTFGASCVT 303
 Db 240 CTGPRSDCLVCQKQDEATCKDTCPLMLNPTTY-----QMDVNEPKGSFATCVK 293
 Qy 304 ACPNYLSTDVSGCTLVCLPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAT 363
 Db 294 KCPRNVVTDHSGCVRCACGPDYEV-EEGIRKCKKCDGCPCKVCNGIGIGISFKDITLSIN 352
 Qy 364 SANTQEPAGCKKIFGSLAFPLSPEDCPDASNTAPLOPQLOQVVFETLEETITGLYLSAWPD 423
 Db 353 ATNKHFKYCTAISGDLHLVPAFKGDSFTPTPLDPRELEILTKVKEITGELLQAWPD 412
 Qy 424 SLPLDSVFONLQVIRGRILHNGAVSLTLQGLGISWLGRLSRLSELGSLALIHHTHLCFV 483
 Db 413 NWTDLHAFENLEIRGRTKHQGFSLAVVGLNITSLGLRSLKEISDGDVIIISGNRLCYA 472
 Qy 484 HTVPMWQLFRPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVNCSQFLRG 543
 Db 473 NTINWKLFGTPNQKTKIMNRAEKDKAVNHVCNPLCSSEGCWGPPEPRDCVSCQVSRG 532
 Qy 544 QECVEECVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFCV 603
 Db 533 RECVEKCNILEGEPRFVENSECICQHPCECLPQAMNITCTGRGPONCICQAHYIDGPHCV 592
 Qy 604 ARCPGKVPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCFAEQASPLTSIVA 663
 Db 593 KTCPAGIMGENTL-VWKYADANNVCHLCHANCTYGCAGPLQGCVEWSPGPKPSIATG 651
 Qy 664 VVGILLVVLGVVFGI-LIKRRQQRKIRYTMRRLLQETLEVEPLTPSGAMPNQAOIRL 722
 Db 652 IVGGLLFIVV-VALGIGLFMRHRHVRKTRLLRLLQERLEVEPLTPSGEAPNQAHILK 710
 Qy 723 ETELRKVKVLSGAGFTYVKGWIPDGNVKIPVAIKVLRNTPSKANKEILDEAYVMAG 782
 Db 711 ETEFKKIKVLSGAGFTYVKGWIPDGNVKIPVAIKVLRNTPSKANKEILDEAYVMAS 770
 Qy 783 VGSVYSRLGLCTSTVOLVTLMPYGLLDHVRNRLGSDLLANCMQIAKGMVYL 842
 Db 771 VDNHVCRLGLCTSTVOLITQLMPYGLLDYVREHKONIGSQVLLNWCQVIAKGMVYL 830
 Qy 843 EDVRLVHRDLAARNVVKSPNHVKITDFGLARLLIDIDETEHADGCKVPIKWMALLESILR 902
 Db 831 EDRLVHRDLAARNVVKSPNHVKITDFGLAKLGAEEKEYHAEGKVPIKWMALLESILH 890
 Qy 903 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMV 962
 Db 891 RIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASDISSILEKGERLPQPPICITIDVYIMV 950
 Qy 963 KCMWIDSCRPRFRELVSERMDARDPQFVVIQ-NEDLGPASPLDSTFYRSLLEDDDMG 1021
 Db 951 KCMWIDASRPKFRLEILIEFQOMARDPQVYVIOGDERMHLPSPTDSNFYRALMDEEDME 1010
 Qy 1022 DLVDAEYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1081
 Db 1011 DVVDAEYLVPQGGFF-----NSPST-----SRTPL 1036
 Qy 1082 APSEGAGSDVFDGLMGAAKGLQSLPHDPSPLQRYSEDPTVPLPSET--DGYVAPLTC 1139
 Db 1037 LSSLATSIN---NSTVACINRNGSCRVEDAFQRYSSDPTGAVTEDIDDAFL----- 1087
 Qy 1140 SPOEYVNPQDVRPOPPSPREGPLPAARPAAGATLERAKTLPCKGVKVDVFAFGAVEN 1199
 Db 1088 -PVPEYVNPQ-SVPKRPAGSVQNVVTHNQPLHP-----APGRDLHYQN--PHSNVGN 1135
 Qy 1200 PEYL-TPQGAAPPHPPAPFSPAFDNLVYWDQ-----DP-----PERCAPPST 1242
 Db 1136 PEYLNTA-----PTCLSSGFSNFPALVQKSHQMSLDNPDYQDFFPKETKPGCI 1186
 Qy 1243 FKGTPTAENPYGLDVP 1260
 Db 1187 FKGTPTAENPYGLDVP 1203

```

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (SC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, P.; Winkler, M.; Ullrich, A.; Vennstr
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:98261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:95228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <BE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <BE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:719-984/Domain: protein kinase ATP-binding motif
F:136-202/Region: protein kinase ATP-binding site: carbohydrate (Thr) (covalent) #
F:192-650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 44.9%; Score 3071.5; DB 1; Length 1223;
Best Local Similarity 48.1%; Pred. No. 6.1e-119;
Matches 627; Conservative 176; Mismatches 349; Indels 151; Gaps 26;
QY 8 RWGLLLALLPPGAA-----STQVCTCTDMKRLPASPETHLDMLRHLXOGCVVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEKVKCOGTNNKLLQLGHVEDHFTSLQWYNNCEVLSNLE 72
QY 62 LTYLPTNASLSFODIQEOVGYYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTFLTKIQEVAGYVLIALNMVDVIPLENLIQIRGNVLYDNSFALAVLSNYH 132
QY 122 PLNNTPTVTGASPGGLREQLRSLTEILAGVLIQNPOLCYODTILWKDIFHNQALAL 181
DB 133 -MNKTO-----GLRELPKMLSEILNGGVKISNNPKLCNMDTILWINDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKGSCWGESSEDQSLTRTVCAAGCA-CKGPLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCOTLTKVCAQOCSCRCGRKVPSDCHNQ 242
QY 240 CAAGCTGPKHSOCLACHFNHSGICELHCPALVFNFTVSFWLRVPKVSASHLERYTGA 299
DB 243 CAAGCTGPRSDCLACRKFDDATCKDTCPLVLNPTTY-----QMDVNPBGKYSFGA 296
QY 300 SCVTACPNYLSFDVGSCTLVCPLNHQEYTAEDGTORCEKSKPCARVCYGLGMEHLREV 359
DB 297 TCURECPHYVVTDHGSCVRSNCNTDTEY-EENGVRKCKCKDGLCSKVCNGIGIGELKGI 355
QY 360 RAVTSANIQEFAGCKKIFGLAFPLPSFDGDPASNTAPLQPEQLQVFETLEETIGLYLIS 419

```

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Flouman, G.D.; Culoucou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderma

```

DB 356 LSNATNIDSFKNCTKINGDVSIIPVAFLDGAFKTKTLPDLPKKLDVFRVKEISGFLDIQ 415
QY 420 AWPDSLPLDSVFQNLQVIRGILHNGAYSLTLQGLIGISWLGRLSRLSGGLALIHNNH 479
DB 416 AWPDNATDLVAFENLEIIRGRTKHQGOYSLAVNVLKTSGLRSLSKESDSDIALMKKNQ 475
QY 480 LCFVHTVPDQOLFNRPHQALHTANRPEDECVGSEGLACHQICARHGCHMGPGPTQCVCSQ 539
DB 476 LCYADTWNRSLEFATQSKTKIIQNRKNKNDCTADRHVCDPLCSVDGCGMGPGPFHCFSCRF 535
QY 540 FLRGQECVECRVLQGLPREVYNARHCLPCHPECPQNG---SVTCFGEADOCVACAHY 596
DB 536 FSROKECVKQCNIILQGEPRFEFERSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHF 595
QY 597 KDPFPCVARCPGKVPDLSYMPIWKFPEDEGACQPCPINCTHSCVDDLDDKGPAPQORASP 656
DB 596 IDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCNTRCGKGLGECGP---NSGK 651
QY 657 LTSIVSAV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQ 715
DB 652 TPSIAGVVGGLLCLVVVGLGIGLYLRR-HIVAKRTLRLRLQRELVLEPLTPSGEAPNQ 710
QY 716 AQMRILKETELRKVKVLGSGARCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLD 775
DB 711 AHLRLKETEFKKVKVVLGSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANKETLD 770
QY 776 EAYVMAGVGPYVSRLLGIGLITSTVQLVTQMLPMYVGLLDHVRNRRGLSGDDLNNWCQI 835
DB 771 EAYVMASVDNPHVCRLLGIGLITSTVQLITQMLPMYVGLLDYIREHKNIGSOYLLNWCQI 830
QY 836 AKMSYLEVDVRLVHRDLAARNVLKSPNNVKITDFGLARLLDIDETEHADGGKVPKIM 895
DB 831 AKGMNYLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADKEKHAEGGVPIKIM 890
QY 896 ALESILRRRTHOSDVMVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTI 955
DB 891 ALESILHRIYTHOSDVMVSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPICTI 950
QY 956 DVYMIWVKMIDSECRPRELVSFSESRMARDPQRFVLIQ-NEDLGPASPLDSTFYRSL 1014
DB 951 DVYMIWVKMIDSECRPRELVSFSESRMARDPQRFVLIQ-NEDLGPASPLDSTFYRSL 1010
QY 1015 LEDDDMDGLVDABEYLVPOQGFPPDPAPGAGGVVHHRSSRSTGGGDLTLGLEPSEE 1074
DB 1011 MEEDMEDIVDABEYLVPHQGF-----NSPST----- 1038
QY 1075 EAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLORYSDPTVPLPSE 1129
DB 1039 --SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQRYSSDPTGNFLEE 1088
QY 1130 T--DGVPALTCSPQPEYVNPQDVRPOPSPREGPLPAARAGATLERAKTLPSPKNGW 1187
DB 1089 SIDDGL-----PAPEYVNO--LMPKKPS-----TAMVQNOIY 1119
QY 1188 KDVF-----AFGAVENPEYLTTPQGAAPQHPHPPAFSPAFDNLNLYMDQ- 1231
DB 1120 NNISLTAISKLPMDRSYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPYIQS 1171
QY 1232 -----DPE-----RGAPSTFKGTPTAENPEYLGDDVP 1260
DB 1172 GNHQINLNDPDYQDDFLPNETKPNGLLKVPAEENPEYLRVAAP 1214

```

A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PJO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AA859446.1; PID:g337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.4%; Score 2971.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 8.2e-115;
Matches 612; Conservative 187; Mismatches 373; Indels 181; Gaps 30;

```
QY 9 WGLLLALLPPGNA-----STQVCTGTDMLKRLPASBETHLDMLRHLYQCGVQGNLELTY 64
DB 8 WWMVSLVAAGTVQPSDSQSCAGTENKLSLSLEQVYRALRYKCYENCEVVMGNLEITS 67
QY 65 LPTNASLSFLDIOIEQVGYVLIHQAHQVQVQLRIRVRGTQLFEDNVALALVDNGDPLN 124
DB 68 IEHNRDLSFLRSRVETGYVLVALNQFYLPLENRIIRGTLYEDRYALALFLNRYKDG 127
QY 125 NTPPTVGASPGGLRELRLSLTEILKGGVLIQRPQLCYQDTILMKDIFHKNNQALTLI 184
DB 128 NF-----GLQSLGKLNLTIELNGGVYVDONKFLCYADTIHQDIVRNWPNSNLTIV 178
QY 185 DTRSRACHPCSPMKGRWCSESSDCQSLTRTVCAGGC-ARCGPLPTDCHCQCQAG 243
DB 179 STNGSSGGRCHKSCGTG-RCMGPTENHCQTLRTVCAEQCGRCGYGVYSDCHRECA 237
QY 244 CTGPKHSDCLACHFNHSGICELHCP-ALVFNNFTVSFWLRVPKVSASHLRYTFGASCV 302
DB 238 CSGPKDTCFACMNDGSGACVTCQPTQFVYNTTF-----QLEHFNKAYITGAFCV 290
QY 303 TACPYNILSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 362
DB 291 KCKPHNFV-VDSSSCVRACSPSKMEV-BENGIMCKPCTDIPCACDGTGTSLSAQTV 348
QY 363 TSANIQEAGCKIKGSLAFLPESPDGPASNTAPLOEOLQVFTLEBITLYLISAWP 422
DB 349 DSSNIDKFINCTKNGNLIFLVTGTHGDPYNAIEAIDPEKLMVFTVREITGFLNIQWP 408
QY 423 DSLPDSYFQNLQVIRGRIHNGAYSLTIQGLIGISWLSRLSRELGSGLALTHHNTLCF 482
DB 409 PNMTDFSVSNLTGVRVLSGLSLLILKQGITSLQFOSKEISAGNIYITDSNLCY 468
QY 483 VHTVPWDOLFNRPHQALLHTANRPEDECVGEGACHQLCARGHCWPGPTQCVCNCSQFLR 542
DB 469 YHTINWTLFTSTINQIRVIRNDRKAENCTAEGMVCNHLCSDDCGWPGPDQCLSKRFSR 528
QY 543 GOECVEECRVLOGLPREYVNAHCLPCHPECP-QNGSVTCRGPADOCVACAHYKDPFP 601
DB 529 GRICIESNLVDGEFREFENGISICEVDPCQCEKMDGLLTCHGPGPDNCTKCSHPKDPN 588
QY 602 CVARCPGKPDLSYMPYKFPDEGACOPCPINCTHSCVLDLDDKGC-----PA 650
DB 589 CVEKCPDGLGANSF--IFKYADPRECHPCNCTGCGNPTSHDCIYYPWTHGSHLTPQ 646
QY 651 BORASPLTSIVASV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPS 709
DB 647 HAR-TPL--IAAGVTGGLFVLVGLTFAVYVRKSIK-KKRALRRFL-ETELVEPLTPS 701
QY 710 GAMPNQAQMTILKETELRKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRWNTSPKA 769
DB 702 GTAPNQAQRLIKETELRKRVVLGSGAGTVYKGIWVPEGETVFKIPVAIKILNETYGPKA 761
QY 770 NKEILDEAVMAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVRNRCRLGSDQLL 829
DB 762 NVEFWEALIMASMDHPHLVRLGLVCLSPFTQLVQLMPHGCLLEYVVEHKDNIGSULL 821
QY 830 NWCWOIAKMGVLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGK 889
```

```
DB 822 NWCVOIAKMGVLEBERRLVHRDLAARNVLKSPNHVKITDFGLARLLGDEKEYNADGGK 881
QY 890 VPIKMMALLESILRRRFTHQSDVWSYGVTTWELMTGAKPYDGIIPAREIPDLLEKGERLPO 949
DB 882 MPIKMMALECITHYRKFTHQSDVWSYGVTTWELMTFGGPKYDGIPTREIPDLLEKGERLPO 941
QY 950 PPICTIDVYIMVVKCMIDSECRPRFRELVSFSSRMARDPQRFVVIQNEED-LGPASPLDS 1008
DB 942 PPICTIDVYIMVVKCMIDADSRRPKFELAAEFSSMARDPQRYLVLIQDDRMKLPSPNDS 1001
QY 1009 TFYRSLLEDDMGDLVDABEYLVPOQGFPCPDPAFGAGGMVHRRSSSTRSGGDLTLTG 1068
DB 1002 KFFQNLLEDELDMDMAEYLVLP-QAFNIPPP-----IYTSRAIDSNRS-----EIG 1049
QY 1069 LEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGNG 1099
DB 1050 HSPPPPAYTPMSGNOFYVYRDGGFAAQGVSVVPYRAPSTIPEAPVAGATAEIFDDSCCNG 1109
QY 1100 AAKGLQSLPTHDPSPLOQYSEDPTVPLPS-----ETDGYVAPLTCSPOPEVYNQPDVR 1152
DB 1110 TLRKPVAVHVOEDSSTQRYSDPTVPAPERSPRGELDEGYMTPMRDKPKQEYLNPEV-- 1167
QY 1153 PQQPSPREGLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGAVENPEYLTPOGGAAPO 1212
DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPVHNASG----- 1194
QY 1213 PHPPPA-----FSPAFDNLVYWDQDPPPERGA 1238
DB 1195 --PPKAEDEYVNEPLYLNTFANTGLKAEYLNILSMPEKAKKAFDNDPDYMNHSLPRST 1252
QY 1239 --PRSTFGTPT-----AENPEYL 1255
DB 1253 LQHPDYLYEYTKYFKQNGRIRPIVAENPEYL 1285
```

RESULT 8

S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homolog; kinase-related transform:
C:Species: Xiphophorus maculatus (southern platyfish)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C:Accession: S06142; S13809
R:Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe:
Nature 341, 415-421, 1989
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu:
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WIT>
A:Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R:Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphoph:
A:Reference number: S13807; MUID:91125882; PMID:1846957
A:Accession: S13809
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, N', 1027-1098, A', 1100-1166 <ADA>
A:Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/3
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; t:
F:1-1166/Domain: signal sequence #status predicted <SIG>
F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-732/Region: protein kinase ATP-binding motif

Query Match 39.1%; Score 2674; DB 1; Length 1166;
Best Local Similarity 45.0%; Pred. No. 1.2e-102;

Db 579 FRGPHCVNSCPHGILG--AKGPIYKYPDAQNECRPCHECTQCNGPELQDCLGQAEVL 636
Qy 654 ASPLTSIVSAVGILLVVLGVVFGILIKRRQOKIR-KYTMRRLLQETELVEPLTPSGAM 712
Db 637 MSKPHLVIAVTVG--LAVILMILGSPFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EK 693
Qy 713 PNAQWRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
Db 694 ANKVLARIPKETELRLKVLGSGVFGTVHKGIIWIPGESIKIPICIKVIEDKSGRQFOA 753
Qy 773 ILDEAYVMAGVSPYVSRLLGICLTSTVQLVLTOLMPYGCLLDHVRENRRGLSGDOLLNC 832
Db 754 VTHMLAVGSLDHAHIVRLGLCPGSSQLVTQVPLGSLLDHVKQHRETLGPOLLNWG 813
Qy 833 MQIAKMSYLEDVRLVRLDRAARNVLKSPNHVKITDFGLARLLIDIDETEHADGKVP 892
Db 814 VOIAKGMWYLEESMVHRDLALRNVLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPI 873
Qy 893 KWALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPI 952
Db 874 KWALESIHFGKYTHQSDVMSYGVTVWELMTFGAEPYAGRLABIPDLLEKGERLAQPOI 933
Qy 953 CTIDVVMYKMWKIDSECRPRFELVSEFSRMARDPQRFVJQNEIDLGPASPLDSTFYR 1012
Db 934 CTIDVVMYKMWKIDSECRPRFELVSEFSRMARDPQRFVJQNEIDLGPASPLDSTFYR 990
Qy 1013 SLEDDDDMGDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRHSSTSGGGDLTLGLEPS 1072
Db 991 SVLTTEL-----QEALELEP-----DLDLLEAE 1016
Qy 1073 EE-----EAPRSLAPSEG-----AGSDVDGDLGMAAKGL 1104
Db 1017 EELGATSLGALSULTGTTLTPRGSSQLLSPSSGYMPMNQSSLCEACLDLAVLGGRRQFS 1076
Qy 1105 QSLPHTDPSPLQRYSEDPTPLPSETDGVV-----APL-----TC-----SPOPE----- 1144
Db 1077 RPISLH-PIPRGR-----PASESEGHVGTGEAELEQKSVCSRSRSRSPRPRGDSA 1128
Qy 1145 YVNPQDVRPPOPPSPREGP-----LPAARPAGATLAKTLSP-GKNGVYV----- 1187
Db 1129 YHSQRSHLLTPVPLSPGLLEEDNGYVMPDTHLRGASSRECTLSSVGLSSVLGTTEE 1188
Qy 1188 -KDVFAFGAGVENPEYLTPOGGAAPQPHPP 1216
Db 1189 DED-----EEYEMNRKRGSP-PRPP 1209

RESULT 11

TVFVLV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif

F;229/Active site: Lys #status predicted

Query Match 25.8%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 1.1e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;
Qy 584 GPBDAQCVACAHYKDPFPFCVARGSPGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 643
Db 60 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPCHRGCKGP 116
Qy 644 DDKCPRAEQRASPLTSIVSAV--GILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETEL 702
Db 117 GLEGCP---NGSTPSPAAAGVWGGLCLVVGIGLYLRRR-HIVKRTLRLLQEREL 172
Qy 703 VEPLTPSGAMPNOAQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVL 762
Db 173 VEPLTPSGEAPNQHLRLKETEFKVKVLOGSGAFGVYKGLWIPGEKVKIPVAIKEUR 232
Qy 763 ENTSPKANKILDEAYVMAGVSPYVSRLLGICLTSTVQLVLTOLMPYGCLLDHVRENRR 822
Db 233 EATSPKANKILDEAYVMASVDNPHVCRLLGICLTSTVQLITOLMPYGCLLDYIREHKN 292
Qy 823 LGSODLLNWCMTAKGHSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLIDETE 882
Db 293 IGSQYLLNWCVOIAKGMVLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGADEKE 352
Qy 883 YHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942
Db 353 YHAEKKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
Qy 943 KGERLPQPPCTIDVYIMVKMWKIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 1001
Db 413 KGERLPQPPCTIDVYIMVKMWKIDADSRKRELIAEFSKMDRPPRYLVIQDERMH 472
Qy 1002 PASPLDSTFYRSLEDDMGDLVDAEYLVPOQGFCCPDPAAGAGGMVHHRHSSTSG 1061
Db 473 LPSPTDSKFYRLTMEEDMEDIVDAEYLVPHQGF-----NSPST----- 513
Qy 1062 GGDLLTGLEPSEEAERPSPL-----APSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSPLQ 1116
Db 514 -----SRTLPLSSLSATSNNSATNCID-----RNGQGHFVREDSESVQ 550
Qy 1117 RYSEDPVPLPSET--DGYVAPLTCSPQPYVNPQDVRPPOPPSPREGPLPAARPAGATLE 1174
Db 551 RYSSDPTGNFLESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
Qy 1175 RAKTLSPGKGVVKKVDF-----APGGAIVENPEYLTPOGGAAPQPHPPAP 1219
Db 586 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1220 SPAPFDNLYWDO-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 634 KTVFESSPYIQSGNHQINLDNPDYQODFLPNETKPNGLLKVPAENPEYLRVAAP 689

RESULT 12

TVYUHV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fan
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaisse, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martir
Science 224, 1456-1459, 1984
A;Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type c
A;Reference number: A38022; MUID:84223957; PMID:6328658

A;Accession: A38022

A;Molecule type: DNA

A;Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DEB>

A;Cross-references: GB:K02006

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific p

F;130-395/Domain: protein kinase homology <KIN>

F;138-146/Region: protein kinase ATP-binding motif

F;165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;

Best Local Similarity 52.2%; Pred. No. 4e-63;

Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 593 CAHYKDPFFCVARCPGKVPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKCPAEQ 652

Db 3 CAHFDGPHCVKACPAAGVGENDTL-VKRYADANAVCOLCHPNCTRGCKGPGLEGCP--- 58

Qy 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGA 711

Db 59 NGSKTPSIAGVVGGLCLLVVGLGILYLRH-HIVKRTLRLLQERELVEPLTPSGE 117

Qy 712 MPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVRENTSPKANK 771

Db 118 APNOAHLRIKETEFKPKVKVLGSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177

Qy 772 EILDEAYVMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGQDLLNW 831

Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVQLITLQMPYGCCLLDYIREHKONIGSYQLLNW 237

Qy 832 CMQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891

Db 238 CVQIAKMNYLEERLVRDLAARNVLKTPQHKVITDFGLAKLLGADEKEYHAEGGKVP 297

Qy 892 IKWMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPP 951

Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLEKGBRLPOPP 357

Qy 952 ICTIDVYIMVKWMIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDSTF 1010

Db 358 ICTIDVYIMVKWMIDADSRPKFRELIAEFSKWARDPPRYLVIQGDERMHLPSPTDSKF 417

Qy 1011 YRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSGGDLTLGLE 1070

Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

Qy 1071 PSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSLPORYSEDPTVP 1125

Db 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFVQRYSSDPTGN 495

Qy 1126 LPSET-DGVVAPLTCSPQPEYVNPQDVRPPOPPSPREGPLPAARPAAGATLERAKTLSPGK 1183

Db 496 FLEESIDGFL-----PAPYVNO--LMPKKPSTAM----- 524

Qy 1184 NGVVKDVFAF-----GGAVENPEYLTQGGAAPOPHPPAPFAPFA 1224

Db 525 --VQNIYNFISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLAKTVFE 574

Qy 1225 NLYWDQDPPRGAPPSTFKTPTAENPEY 1254

Db 575 SSPYMIQSGNHQ-----INLDNPDY 594

RESULT 13

S35745

Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997

C;Accession: S35745

R;Vennstrom, B.

submitted to the EMBL Data Library, March 1993

A;Reference number: S35743

A;Accession: S35745

A;Molecule type: DNA

A;Residues: 1-544 <VEN>

A;Cross-references: EMBL:X12707

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

F;170/Active site: Lys #status predicted

Query Match 24.1%; Score 1647; DB 2; Length 544;

Best Local Similarity 54.9%; Pred. No. 7.1e-61;

Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

Qy 584 GPEADQCACAHYKDPFFCVARCPGKVPDLSPYIWKFPDEEGACQPCPINCTHSCVDL 643

Db 1 GP--DHCMKCAHFTDGPCHVCACPAAGVGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 57

Qy 644 DDGCPAEQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 702

Db 58 GLEGCP---NGSKTPSIAGVVGGLCLLVVGLGILYLRH-HIVKRTLRLLQEREL 113

Qy 703 VEPLTPSGAMPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR 762

Db 114 VEPLTPSGEAPNOAHLRIKETEFKPKVKVLGSGAFGVYKGLWIPEGEKVTIPVAIKELR 173

Qy 763 ENTSPKANKIILDEAYVMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRR 822

Db 174 EATSPKANKIILDEAYVMASVDNPHVCRLLGICLTSTVQLITLQMPYGCCLLDYIREHKDN 233

Qy 823 LGSODLLNCWQIAKMSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882

Db 234 IGSYLLNWCVCQIAKMNYLEERLVRDLAARNVLKTPQHKVITDFGLAKQLGADBEKE 293

Qy 883 YHADGGKVPKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 942

Db 294 YHAEGGKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISVLE 353

Qy 943 KGBRLPOPPICTIDVYIMVKWMIDSECRPRELVSFSEMRARDPQRFVVIQ-NEDLG 1001

Db 354 KGBRLPOPPICTIDVYIMVKWMSDADSRPKFRELIAEFSKWARDPPRYLVIQGDERMH 413

Qy 1002 PASPLDSTFVRSILLEDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSG 1061

Db 414 LPSPTDSKPYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

Qy 1062 GGDITLGLPSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQ 1116

Db 455 -----SRTPLLSLSATSNNSATNCIDRNGG-----H----- 481

Qy 1117 RYSEDPTVLPSETDGVVAPLTCSPQPEYVNPQDVRPPOPPSPREGPLPAARPAAGAT-LER 1175

Db 482 -----PVREDGFL-----PAPYVNO--LMPKKPSTAMQNIYVVISLTAISK 523

Qy 1176 AKTILSPKNGVKKDVFAPFGGAVENPEYL 1203

Db 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C;Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus muti

A;Reference number: S00727; MUID:88217326; PMID:2897102

A;Accession: S00727

A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 24.0%; Score 1640; DB 2; Length 545;
Beat Local Similarity 54.9%; Pred. No. 1.4e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 584 GPEADQCACAHYKDPDFCFVACRCSGVKPKDLSYMPIWKFPDEGACQPCINCTHSCVDL 643
Db 1 GP--DHCWKAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCQLCHPNCCTRGCKGP 57
Qy 644 DDKCPAEQASPLTISVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETEL 703
Db 58 GLEGP---NGSKTPSAAGVVGGLLCLVVGGLGIGLYLRRR-HIVKRTLRLLQEREL 113
Qy 703 VEPLTPSGAMPNOAOMRILKETELRKVKVLGSGAGFTVYGIWIPDGENKVIPIAVKLR 762
Db 114 VEPULTPSGEAPNQAHRLILKETEPKKVKLGFAGFTVYKGLWIPEGEKVTPIPVAIKELR 173
Qy 763 ENTSPKANKEILDYAYVMAGVSGSPVSRLLGICLTSTVQLVTQLMPYGCILLDHYRENRGR 822
Db 174 EATSPKANKEILDYAYVMASVDNPHVCELLGICLTSTVQLITQLMPYGCILLDYIREHKN 233
Qy 823 LGSODLLNWCQIAKGSYSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 234 IGSQYLLNWCQIAKGMNYLEERHLVHRDLAARNVLKTPDVKITDFGLAKQLGADEKE 293
Qy 883 YHADGGKVPKIMWALESILRRFTHQSDVMSYGYVWELMTFGAKPDGIPAREIDPILLE 942
Db 294 YHAEKKVPKIMWALESILHRIYTHQSDVMSYGYVWELMTFGSKPDGIPASEISSVLE 353
Qy 943 KGERLPOPPICTIDVYIMVKWMIIDSECPFRFRELVSFSESMARDPQRFVVIQ-NSDLG 1001
Db 354 KGERLPOPPICTIDVYIMVKWMSDADSRPKFRELIAEFSKWARDPPRVLVITQGDGRMH 413
Qy 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDPAAGGMVHHRHSSSTRSG 1061
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454
Qy 1062 GGDLTLCLEPSEEEAPRSP-----ASSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQ 1116
Db 455 -----SRTLSSLASATSNNSATNCIDRNG-----H-----481
Qy 1117 RYSEDPVPLPSETDGVVAPLTCSPQPEYVNOPDVRPQPSREGPLPAARPAGAT-LER 1175
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVYISLTAISK 523
Qy 1176 AKTLSPGKNGVKVDVFAFGAVENPEYL 1203
Db 524 LPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15
B44776
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
C:Species: avian erythroblastosis virus
C:Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
C:Accession: B44776
R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
Oncogene 5, 15-24, 1990
A:Title: Six amino acids from the retroviral gene gag greatly enhance the trans-
A:Reference number: A44776; MUID:90206603; PMID:1969616
A:Accession: B44776
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-540 <BRU>

RESULT 15
 B44776
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)
 C:Species: avian erythroblastosis virus
 C:date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000
 C:Accession: B44776
 R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCarley, D.J.; Schatzman, R.C.
 Oncogene 5, 15-24, 1990
 A:title: Six amino acids from the retroviral gene gag greatly enhance the transforming
 A:Reference number: A44776; MUID:90206603; PMID:1969616
 A:Accession: B44776
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-540 <BRU>

A;Cross-references: GB:X52211
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor tyrosine kinase
C;Keywords: ATP; phosphotransferase
F;130-395/Domain: protein kinase
F;138-146/Region: protein kinase
Query Match Best Local Similarity 54. Matches 340; Conservative
Qy 593 CAHYKDPFCVACPP
Dd 3 CAHFDGPHCKACPP
Qy 653 RASPLTSTVSAAV-G G
Dd 59 NGSKTPSIAGVGG
Qy 712 WPNQAQRILKETEL
Dd 118 APNOAHLRLKETEF
Qy 772 ELDEAYVMAGVGSP
Dd 178 ELDEAYMASVDNPI
Qy 832 CMQIAKGSYLEDVRR
Dd 238 CVQIAKMNYLEERH
Qy 892 IKWMALESILRRPFT
Dd 298 IKWMALESILRIYTT
Qy 952 ICTIDVTMIMVKWM
Dd 358 ICTIDVTMIMVKWM
Qy 1011 YRSLLDDMDGLVDV
Dd 418 YRTLMEEDMEDVD
Qy 1071 PSEEAARPSPL
Dd 450 -----SRTELLSSLS
Qy 1126 LPSETDGYAPLTCS
Dd 477 -PVREDGF-----F
Qy 1185 GWKKDFAFGGAVEN
Dd 528 -----SHSTAVDN

Search completed: July 22, 2003, 09:09:15
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds
(without alignment)
5088.033 Million cell updates/sec

Title: SEQ4-250-264-14

Perfect score: 6839

Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6702	98.0	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5901	86.3	1257	1 ERB2_RAT	P06494 rattus norv
3	5880.5	86.0	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3114	45.0	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3093	43.2	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2971.5	43.4	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2952	43.2	1308	1 ERB4_RAT	Q62956 rattus norv
8	2690.5	39.3	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2400.5	35.1	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2329.5	34.1	1339	1 ERB3_RAT	Q62799 rattus norv
11	1922	28.1	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.6	634	1 ERBB_ALV	P00534 avian leuko
13	1703	24.9	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.8	540	1 ERBB_AVIEU	P11273 avian eryth
15	1572	23.0	703	1 EGFR_CHICK	P13387 gallus gall
16	1265	18.5	1323	1 L723_CAEPL	P24348 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70424 mus musculu
18	733	10.7	1363	1 ILPR_ERALA	O02466 branchiosto
19	721.5	10.5	1372	1 INSR_MOUSE	P15208 mus musculu
20	719	10.5	1383	1 INSR_RAT	P15127 rattus norv
21	716	10.5	1382	1 INSR_HUMAN	P06213 homo sapien
22	699	10.2	1300	1 IRR_MOUSE	Q9wt14 mus musculu
23	693	10.1	1297	1 IRR_HUMAN	P4616 homo sapien
24	691	10.1	1607	1 MIPR_LYMT	Q25410 lymphocita
25	687	10.0	1477	1 ITRK_HYDAT	Q25197 hydra atten
26	686.5	10.0	1300	1 IRR_CAVPO	P14617 cavia porce
27	649	9.5	1367	1 IGR1_HUMAN	P08069 homo sapien
28	636	9.3	1373	1 IGR1_MOUSE	Q60751 mus musculu
29	632.5	9.2	1370	1 IGR1_RAT	P24062 rattus norv
30	620	9.1	1390	1 INSR_AEDAE	Q9j105 aedes aegypt
31	606	8.9	2146	1 INSR_DROME	P09208 drosophila
32	594.5	8.7	987	1 EPB4_HUMAN	P54760 homo sapien
33	589.5	8.6	977	1 EPB2_MOUSE	Q03145 mus musculu

ALIGNMENTS

RESULT 1

ID	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.
AC	P04626;			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)			
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).			
GN	ERBB2 OR HER2 OR NGL OR NEU.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86118663; PubMed=3003577;			
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,			
RA	Saito T., Toyoshima K.;			
RT	"Similarity of protein encoded by the human c-erbB-2 gene to			
RT	epidermal growth factor receptor.";			
RL	Nature 319:230-234(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86070181; PubMed=2999974;			
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,			
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,			
RA	Frankke U., Levinson A., Ullrich A.;			
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor			
RT	shares chromosomal location with neu oncogene.";			
RL	Science 230:1132-1139(1985).			
RN	[3]			
RP	SEQUENCE OF 737-1031 FROM N.A.			
RX	MEDLINE=86016729; PubMed=2995967;			
RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;			
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the			
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a			
RT	human salivary gland adenocarcinoma.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).			
RN	[4]			
RP	VARIANTS VAL-654 AND VAL-655.			
RX	MEDLINE=93194196; PubMed=8095488;			
RA	Eheanl A., Low J., Wallace R.B., Wu A.M.;			
RT	"Characterization of a new allele of the human ERBB2 gene by allele-			
RT	specific competition hybridization.";			
RL	Genomics 15:426-429(1993).			
CC	- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,			
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A			
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-			
CC	ALPHA AND AMPHIREGULIN.			
CC	- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein			
CC	tyrosine phosphate.			
CC	- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS			
CC	(POTENTIAL).			
CC	- SUBCELLULAR LOCATION: Type I membrane protein.			

P07949 homo sapien
P29317 homo sapien
Q07494 gallus gall
P54761 mus musculu
P09759 rattus norv
Q00944 gallus gall
Q91736 xenopus lae
Q91571 xenopus lae
Q91738 xenopus lae
P54762 homo sapien
P53356 hydra atten
P34152 mus musculu


```
Qy 721 LKETELRKVKVLGSGAGTGYVKGWIWPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVM 780
Db 715 LKETELRKVKVLGSGAGTGYVKGWIWPDGENVKIPVAIKVLRENTSPKANKEILLDEAYVM 774
Qy 781 AGVSPVSVRLGLCLSTVOLVQLMPYGCGLLDHVRNRCGLSDLLNWCMOIAKMS 840
Db 775 AGVSPVSVRLGLCLSTVOLVQLMPYGCGLLDHVRNRCGLSDLLNWCMOIAKMS 834
Qy 841 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESI 900
Db 835 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESI 894
Qy 901 LRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRELVSFSESMARDPQRFVVIQNEIDLGPAFLDSTFYRSLLDDDM 1020
Db 955 MVKCMWIDSECRPRELVSFSESMARDPQRFVVIQNEIDLGPAFLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQFFCPDPAAGAGGVHRRHSSSTRSGGDLTLGLEPSEEEAARSP 1080
Db 1015 GDLVDAEYLVPOQFFCPDPAAGAGGVHRRHSSSTRSGGDLTLGLEPSEEEAARSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 2
ID_ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
DE ERBB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
```

```
[3]
RN STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Loftis F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RT oncogenic and oncogenic forms of the neu protein.";
RL EMBO J. 11:43-48(1992).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL; X03362; CAA27059.1; ALT_INIT.
PIR; A24562; TVRTNU.
HSSP; P11362; 1FGK.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR004019; YLP_motif.
Pfam; PF00069; pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
Pfam; PF02757; YLP_2; 1.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SM00219; TyKc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
```



```

FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD.RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD.RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.3%; Score 5901; DB 1; Length 1257;
Best Local Similarity 86.5%; Pred. No. 8e-310;
Matches 1092; Conservative 52; Mismatches 111; Indels 8; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTGCTDMKRLRASPETHLDMLRLHYQCCVQGNL 60
Db 1 MELAALCRWGLLLALLPPGIAGTCTGCTDMKRLRASPETHLDMLRLHYQCCVQGNL 60

Qy 61 ELTYLPTNASLFDIOIEVOGYVLIHNOVROVPLQRLRIVRGTLFEDNYALAVLNG 120
Db 61 ELTYVPANASLUSLFDIOIEVOGYMLIAHQVRLQRLRIVRGTLFEDNYALAVLNG 120

Qy 121 DPLNNTTPTV-GASPGGLRELQLSLTEILKGGVLIORNPOLCYODTILWKDIFHNQOL 179
Db 121 DPQDVAASTGRTPEGLRELQLSLTEILKGGVLIORNPQLCYQDMVLWKDVFKNQOL 180

Qy 180 ALTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLRTVTCAGGCARCKGPLPTDCHEQ 239
Db 181 APVDIDNRSRACPPACPKDNHCWGESPEDCQILTGTICTSGCARCKGRLPTDCHEQ 240

Qy 240 CAAGCTGPKHSDCLACLFHNSGTCCLHCPALVFNFTVFWLVRVPKVASHLRYTGA 299
Db 241 CAAGCTGPKHSDCLACLFHNSGTCCLHCPALVNTDTFMSMPE-----GRYTFGA 294

Qy 300 SCVTACPNYLSTDVSGCTLVCPHNGEVTAEDGTQRCCKSKPCARVCYGLGMEHLREV 359
Db 295 SCVTTCPNYLSTEVSGCTLVCPNPNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLRGA 354

Qy 360 RAVTSANTQEFAGCKKIFGSLAFIPESDGPDPASNTAPLQEQLOVFTLEITGYLYIS 419
Db 355 RAITSNDNVQEPDGCKKIFGSLAFIPESDGPDPSSGIAPLRPEQLQVFTLEITGYLYIS 414

Qy 420 AWPDSLPLSVFQNLQVIRGRILHNGAYSLLTQGLISWLGRLSRLRELGSGLALIHNNTH 479
Db 415 AWPDSLRLSVFQNLRIIRGRILHNGAYSLLTQGLIHSLGRSRLRELGSGLALIHNNH 474

Qy 480 LCFVHTVPWDQLFRNPHQALLHTANRPEDE-CVSGGLACHQLCARGHCWGPGPTQCVNCS 538
Db 475 LCFVHTVPWDQLFRNPHQALLHSGNRPEEDLCVSSGLVCNLSIAGHCWGPGPTQCVNCS 534

Qy 539 QFLRGQECVSCRVLQGLIPRYVNAHCLCPCHPECPQNGSVTCFPGPADOCVCAHYKD 598
Db 535 HFLRGQECVSCRVMKGLPREYVSDKRLCLPCHPECPQNSSETCFGSEADQCAACHYKD 594

Qy 599 PPFCVAPCSVKPDLSPMPLWKPPDEGACQPCPINCTHSCVDLDDKGCPEAQRASPLT 658
Db 595 SSSCVAPCSVKPDLSPMPLWKPPDEEGICQPCPINCTHSCVDLDDKGCPEAQRASVPT 654

Qy 659 SIVSAVVGILLVWVVGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAO 718
Db 655 FIATVVGILLVWVVGILIKRRQOKIRKYMRLLOETELVEPLTPSGAMPNQAO 714

Qy 719 RILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAY 778

```

```

Db 715 RILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAY 774
Qy 779 VMAGVSPYVSRLLIGICLTSTVOLVQLMPYGCCLLDHVRENRRGRSGODLLNMCQIANG 838
Db 775 VMAGVSPYVSRLLIGICLTSTVOLVQLMPYGCCLLDHVREHRRGRSGODLLNMCQIANG 834
Qy 839 MSYLEDVRLVHRDLAARNVVKSPNHVKITDIFGLARLLDIDETEHADGCKVPKIMMALE 898
Db 835 MSYLEDVRLVHRDLAARNVVKSPNHVKITDIFGLARLLDIDETEHADGCKVPKIMMALE 894
Qy 899 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVY 958
Db 895 SILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVY 954
Qy 959 MIMVKCMIDSECRPRRELSEFSRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLD 1018
Db 955 MIMVKCMIDSECRPRRELSEFSRWARDQRFVVIQNEIDLGPSPMDSTFYRSLLD 1014
Qy 1019 DMGDLVDAEYLVPOQGFPCDPAPGAGMVRHRRSSSTRSGGDLTLGLPSEEEAPR 1078
Db 1015 DMGDLVDAEYLVPOQGFPCDPAPGAGMVRHRRSSSTRSGGDLTLGLPSEEGP 1074
Qy 1079 SPLAPSEGAGSDVDGDLGMAAKGLQLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLT 1138
Db 1075 SPLAPSEGAGSDVDGDLGMAKGLQLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLA 1134
Qy 1139 CSPOPEYVQDVPVRPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVDYPAFGGAVE 1198
Db 1135 CSPOPEYVQDVPVRPPSPREGPLPAARPAAGATLERAKTILSPGKGVVVDYPAFGGAVE 1194
Qy 1199 NPEYLTPOGGAAPPHPPAFSPAFDNLVYDQDPPERPAPSTFKGTPAENPEYVLGD 1258
Db 1195 NPEYLVPRGTASTPHSPAFSPAFDNLVYDQDPPERPAPSTFKGTPAENPEYVLGD 1254
Qy 1259 VPV 1261
Db 1255 VPV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
GN (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RA MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

```

```

CC CC RESIDUES.
CC CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D16295; BAA03801.1; -
CC CC HSP; p11362; IFGK.
CC CC InterPro; IPR000494; EGFR_L_domain.
CC CC InterPro; IPR000719; Euk_kinase.
CC CC InterPro; IPR002174; Furin-like.
CC CC InterPro; IPR001245; Tyr_kinase.
CC CC InterPro; IPR004019; YLP_motif.
CC CC Pfam; PF00069; kinase; 1.
CC CC Pfam; PF00757; Furin-like; 1.
CC CC Pfam; PF01030; Recep_L_domain; 2.
CC CC Pfam; PF02757; YLP; 2.
CC CC ProDom; PD000001; Euk_kinase; 1.
CC CC SMART; SM00261; FU; 3.
CC CC SMART; SM00219; TyrKc; 1.
CC CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC CC Proto-oncogene; Disease mutation.
CC CC SIGNAL 1 21
CC CC CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
CC CC DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
CC CC TRANSMEM 653 675 POTENTIAL.
CC CC DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
CC CC DOMAIN 158 368 CYS-RICH.
CC CC DOMAIN 472 644 CYS-RICH.
CC CC DOMAIN 720 987 PROTEIN KINASE.
CC CC NP_BIND 726 734 ATP (BY SIMILARITY).
CC CC BINDING 753 753 ATP (BY SIMILARITY).
CC CC ACT_SITE 845 845 BY SIMILARITY.
CC CC DISULFID 195 204 BY SIMILARITY.
CC CC DISULFID 199 212 BY SIMILARITY.
CC CC DISULFID 236 244 BY SIMILARITY.
CC CC DISULFID 240 252 BY SIMILARITY.
CC CC DISULFID 255 264 BY SIMILARITY.
CC CC DISULFID 268 295 BY SIMILARITY.
CC CC DISULFID 299 311 BY SIMILARITY.
CC CC DISULFID 315 331 BY SIMILARITY.
CC CC DISULFID 334 338 BY SIMILARITY.
CC CC DISULFID 511 520 BY SIMILARITY.
CC CC DISULFID 515 528 BY SIMILARITY.
CC CC DISULFID 531 540 BY SIMILARITY.
CC CC DISULFID 544 560 BY SIMILARITY.
CC CC DISULFID 563 576 BY SIMILARITY.
CC CC DISULFID 567 584 BY SIMILARITY.
CC CC DISULFID 587 596 BY SIMILARITY.
CC CC DISULFID 600 623 BY SIMILARITY.
CC CC DISULFID 626 634 BY SIMILARITY.
CC CC DISULFID 630 642 BY SIMILARITY.
CC CC MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC CC MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC CC CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CC VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
CC CC VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
CC CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

```

```

Query Match 86.0%; Score 5880.5; DB 1; Length 1254;
Best Local Similarity 86.1%; Pred. No. 1e-308;
Matches 1086; Conservative 60; Mismatches 108; Indels 7; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASQVCTGTDMLRLPASPEHLDMLRHLYQCGVQVQNL 60
Db 1 MELAAACWGLLLALLSPGASQVCTGTDMLRLPASPEHLDMLRHLYQCGVQVQNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNOVQVPLRLRIVRGTOLFDNVALAVLDNG 120
Db 61 ELTYLPANATLSFLQDIOEVQGYMLIAHQVHVRPLRLRIVRGTOLFDKVALAVLDNR 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFHKKNQLA 180
Db 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIQRPQLCYQDTILWKDIFVRKNQLA 180
Qy 181 LTLIDTNRSRACHPCSPCKSGRCHWSESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
Db 181 PVDIDTNRSRACPPCAPACKDNHCWASPEDCQTLTGTIAPRAVPAARLARLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSWFLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPPE-----GRYTFGAS 294
Qy 301 CVTACPNYLSLTDVSGSLTLCVPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTTCPNYLSLTDVSGSLTLCVPLNNQEVTAEDGTORCEKSKCARVCYGLGMEHLRGAR 354
Qy 361 AVTSANIOEPAGCKKIFGSLAPLPSFDGDPASNTAPIQPEQLQVFETLEEITGYLYISA 420
Db 355 AITSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPIQPEQLQVFETLEEITGYLYISA 414
Qy 421 WPSLSLPLDSVFNQVIRGRVILHNGAYSILTOGLGSLWLSRLSRLSGLSLALIHNTHL 480
Db 415 WPSLHLSLVFNQVIRGRVILHNGAYSILTOGLGSLWLSRLSRLSGLSLALIHNTHL 474
Qy 481 CFVHTVPMWDLFRNPQALLHTANRPEDECVGEGELACHOLCARGHGWGPQTQVNCVCSOF 540
Db 475 CFVHTVPMWDLFRNPQALLHSGNPSEECGLKDFACYPLCAHGHGWGPQTQVNCVCSHF 534
Qy 541 LRQECVVEECVLOGLPREYVNAHCLPCHPECPQNGSVTCFPGPEADQVCAHAHYKDP 600
Db 535 LRQECVKECVKWKGLPREYVNGKHLCPCHPECPQNSTETCTGSEADQCTACPHYKDP 594
Qy 601 FCVARCPSGVKPDLSYMPIMKFPDEGACQCPINCHSCVDLDDKGCFAEQRASPLTSI 660
Db 595 FCVARCPSGVKPDLSYMPIMKFPDEGACQCPINCHSCVDLDERGCFAEQRASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOQMRI 720
Db 655 IATVVGILLVIVGVVGLIKRRROKIRKVTMRRLLOETELVEPLTPSGAMPNOQMRI 714
Qy 721 LKETELRKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 774
Qy 781 AGVGSPPYVRLGLCLTSTVOLVTLQMPYGLLDHVRNRRGLSGODLLNWCQAKGMS 840
Db 775 AGLGSPYVRLGLCLTSTVOLVTLQMPYGLLDHVRNRRGLSGODLLNWCQAKGMS 834
Qy 841 YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI 900
Db 835 YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESI 894
Qy 901 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPPICTIDVYMI 960
Db 895 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGBRLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFTFRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFTFRSLLEDDDM 1014

```

OY 1021 GDLVDAEYLVPPQGFPCDPAPGAGMVGHHRRSSSTRSGGDLTLGLPSEEEAPRSP 1080
|||||
DB 1015 GDLVDAEYLVPPQGFPCDPAPGAGMVGHHRRSSSTRSGGDLTLGLPSEEEAPRSP 1074
|||||
OY 1081 LAPSEAGSDVPDGLGKAAGLQSLPHTHDSPLQRYSEDTVPPLPSETDGYVAPLTCS 1140
|||||
DB 1075 LAPSEAGSDVPDGLGKAAGLQSLPHTHDSPLQRYSEDTVPPLPSETDGYVAPLTCS 1134
|||||
OY 1141 POPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVDVAFGAVENP 1200
|||||
DB 1135 POPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTLSPKNGVVDVAFGAVENP 1194
|||||
OY 1201 EYLTPQGAAPQPPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
|||||
DB 1195 EYLTPQGAAPQPPPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVP 1253
|||||
OY 1261 V 1261
DB 1254 V 1254

RESULT 4
EGFR_HUMAN
ID EGFR_HUMAN STANDARD; PRT: 1210 AA.
AC P00533; P06268; Q14225; Q9UM07; Q9UM08; Q9UMG5; Q92795; Q00732;
AC O00688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
DT 21-JUN-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
protein-tyrosine kinase ErbB-1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta.";
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor.";
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGR) in ovarian cancer.";
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms.";
RL Genomics 71:1-20(2001).
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245935; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor


```
Oy 965 WMIDECRRPRELVSFMRWDPQRFVWTO-NEDLGPAASPLDSTFYRSILLEDDMGDL 1023
Db 951 WMIDADSPKRELIIIEFSKMDPQRYLVITQGDHMLPSTDSNFYRALMDEEDMDV 1010
Oy 1024 VDAEYLVPQGFPCPDPAAGAGVHRRSSSTRSGGDLTGLSESEAEPSPLAP 1083
Db 1011 VDAEYLVPQGFPCPDPAAGAGVHRRSSSTRSGGDLTGLSESEAEPSPLAP 1036
Oy 1084 SEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDTPVLPSET--DGVPAPLTCS 1141
Db 1037 SLSATSN--NSTVACIDRNGLOSCPIKEDSFLQRYSSDPTGALTEDSIDDTFL-----P 1088
Oy 1142 QPEVYNQDVRQPPSPREGPLPAPRAGATLERAKTLSPKNGVGVDFAGGAVENPE 1201
Db 1089 VPEYINQ--SVKRPAGSVQNVYHNPQPLNP-----APSRDHYQD--PHSTAVGNPE 1137
Oy 1202 YL-TPOGAAQPPHPPAFSAFNLVYWDQ-----DP-----PERGAPPSTFK 1244
Db 1138 YLNTVQ-----PTCVNSTFDSPAHMAQKGSQHSILDNPVQDDFFPKAKPNGIFK 1188
Oy 1245 GTPTAENPEYL 1255
Db 1189 GS-TAENAEYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC 001279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
(bek/KGFR) gene."
RL Oncogene 7:1957-1962(1992).
(2)
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation."
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
(3)
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX Hibbs M.L.;
RA Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
(4)
RN SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase."
RL Genes Dev. 8:399-413(1994).
(5)
RN SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
```

```
RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site."
RL Oncogene 6:673-676(1991).
(6)
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC !- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- MISCELLANEOUS: Binding of EGF to the receptor leads to
dimerization, internalization of the EGF-receptor complex,
induction of the tyrosine kinase activity, stimulation of cell DNA
synthesis, and cell proliferation.
CC !- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
EMBL; X78987; CAA55587.1; -
EMBL; U03425; AAA17899.1; -
EMBL; X59698; CAA42219.1; -
EMBL; L06864; AAA53029.1; -
EMBL; Z12608; CAA78249.1; -
HSSP; P11362; IFGK.
MGP; MGI:95294; Egfr.
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
Pfam; PF00069; Pkinase; 1.
Pfam; PF00757; Furin-like; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 3.
SMART; SM00219; TyrcK; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
SIGNAL 1 24
FT CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.
FT DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
FT REPEAT 75 300 APPROXIMATE.
FT REPEAT 390 600 APPROXIMATE.
FT DOMAIN 1028 1071 SER-RICH.
FT DOMAIN 714 981 PROTEIN KINASE.
FT NP_BIND 720 728 ATP (BY SIMILARITY).
FT BINDING 747 747 ATP (BY SIMILARITY).
FT ACT_SITE 839 839 BY SIMILARITY.
FT DISULFID 190 199 BY SIMILARITY.
FT DISULFID 194 207 BY SIMILARITY.
FT DISULFID 215 223 BY SIMILARITY.
FT DISULFID 219 231 BY SIMILARITY.
FT DISULFID 232 240 BY SIMILARITY.
FT DISULFID 236 248 BY SIMILARITY.
FT DISULFID 251 260 BY SIMILARITY.
FT DISULFID 264 291 BY SIMILARITY.
FT DISULFID 295 307 BY SIMILARITY.
FT DISULFID 311 326 BY SIMILARITY.
FT DISULFID 329 333 BY SIMILARITY.
```

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
FT SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 45.2%; Score 3093; DB 1; Length 1210;
Best Local Similarity 49.1%; Pred. No. 6e-159;
Matches 628; Conservative 171; Mismatches 363; Indels 116; Gaps 24;

Qy 11 LLLALLPPGA--STVCTGTGDKMLRLPASBETHLDMLRHLYQCGVQVQGNLEITYLPTN 68
Dy 14 LTLAALCAAGGALEKKVCCQSTNSRLTQLGTGFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLQDIQBVQVGLIAHNOVQLQRLIRVGTOLFEDNYALVALDNGDPLNNTTP 128
Dy 74 YLSFLKTIQVAGVLIANTVERIPLENQIIRGNALYENTYALATLSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPLQCVQDTILWKDI----FHKNNOLALTLI 184
Dy 125 -YGTNRGTGLRELPMNLQELIGAVRFSNNPILCNMDTIQWRDIQVNVFMSNMSDL--- 180
Qy 185 DTRNSRACHPCSPMKCSRGCSGESSEDCQSILTRTVACGGCA-RCKGPLPTDCHEQCAAG 243
Dy 181 -QSHPSKCPKCDPSCPNCSGCGGGEENGCKLTKIACAQCSHRCGRSPSDCHNQCAAG 239
Qy 244 CTGPKHSDCLACHFNHSGICELHCPALVFNFTVSWFLRVPKVSASHLERYTFGACVT 303
Dy 240 CTGPRESDCLVCKQFQDEATCKDTCPPLMLNPTTY-----QMDVNPGEKYSFGATCVK 293
Qy 304 ACPVNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVT 363
Dy 294 KCPRNVTVDHSGSVRACGPDYEV-EEDGIRKCKCDGPKCKVNGIGIGEFKDTLSIN 352
Qy 364 SANIQEFAGCKIIFGSLAFIPESFDGDPASNTAPLOPELOVFTLBEITYLISAWPD 423
Dy 353 ATNIKHFKYCTAISGDLHILPVAFFGDSFTRTPLDPRELEIKTVKEITGFLLIQAWPD 412
Qy 424 SLPDLVSFQNLQVIRGLIHNGVSLTLOGISWLGRLSRLRELGLALIHHTHLFCV 483
Dy 413 NWTDLHAFENLEIRGTQKQHQQLQSLAVVGNLITSLGLRSLKEISDGDVIISGNRNLCA 472
Qy 484 HTVPWDQLFRNPQALHTANRPEDECVGEGLACHQLCARHCWGPQFTQCVNCSQFLRG 543
Dy 473 NTINWKKLFGTPNQKTKIMNRAEKCKAVNHVNCPLCSGCGHGPFRDCVSCQNVSRG 532
Qy 544 QECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHYKDPFPVCV 603

```

```

Db 533 RECVEKCNILEGEPRFVENSECICHPECLPQAMNITCTGRGPNDCIQCAHYIDGPHCV 592
Qy 604 ARCSGVKPDLSYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGCAPORASPLTSIVA 663
Db 593 KTCBPAGINGENNTL--VMKYADANNVCHLCHANCTYGCAGPGLOGCEVWSPGPKIPSIATG 651
Qy 664 VVGILLVVVLGVVFGI-LIKERQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAORILK 722
Db 652 IVGGLLFIVV-VALGIGLFMRHRRHVKTLLRRLLOERELVEPLTPSGEAPNQAHLRLK 710
Qy 723 ETELKVKVLGSGAFGTYYKGIWIPDGENKIPVAIKVLRNTPSKANKEILDEAYVMAG 782
Db 711 ETEFKKIVLGSAGFTYVKGWIPEGEKVIPVAIKELREATSPKANKEILDEAYVMAS 770
Qy 783 VGSPPVSRLLGICLTSTVOLVTLMPYGCGLDHDVRENRGLSGODLLNWCMOIAKMSYL 842
Db 771 VNDPHVCRLLGLCTSTVOLITQLMPYGCGLDHYVREHKDNGISQYLLNWCVOIAKGMNLY 830
Qy 843 EDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIKWMALESILR 902
Db 831 EDRRLVHRDLAARNVVKTPQHVKITDFGLAKLLGAEEKYHAEKGKVPKIKWMALESILH 890
Qy 903 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 962
Db 891 RIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASDISSILEKGERLPQPPICTIDVYIMV 950
Qy 963 KCMWIDSECRFRRELVSERPMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDMG 1021
Db 951 KCMWIDASRPFRELILEFSKMARDPQRYLVIQGDERMHLPSPTDSNFYRALMDEEDE 1010
Qy 1022 DLVDAEYLVPOQGFCCPDPAAGAGVMVHRRSSSTRSGGDLTLGLLEPSEEAAPRSP 1081
Db 1011 DTVDAEYLVPOQGF-----NSPST-----SRTPL 1036
Qy 1082 APSEGAGSDVDGDLGMAKGLQSLTHDPSPLOQYSEDPTVLPSET--DGYVAPLTC 1139
Db 1037 LSSLSATSN---NSTVACINRNGSCRKVEDAFLOQYSSDPTGAVTEDNIDDAFL---- 1087
Qy 1140 SPOEYVYNQPVRRQPPSPREGPLPAAPACATLERAKTLSPKNGVGVKDVFAFGAVEN 1199
Db 1088 -PVPEYVYNQ-SVPKRPAGSVQNVYHNQPLHP-----APGRDLHYQN--PHSNVAVGN 1135
Qy 1200 PEYL-TPQGGAAPHPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPPST 1242
Db 1136 PEYLNTAQ-----PTCLSSGFNSPALMIQKSHQMSLDNDPDYQQDFPFKEPKNGI 1186
Qy 1243 PKGTPTAENPEYLGLDVP 1260
Db 1187 FKQ-PTAENAEYLRVAPP 1203

RESULT 6
ID ERB4 HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

```


epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RP TISSUE-Fetal brain;
 RC MEDLINE=97476287; PubMed=9334263;
 RX Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
 RA Klugebrun M.;
 RA "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC !- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC !- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC !- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC !- SUBCELLULAR LOCATION: Type I membrane protein.
 CC !- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC !- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC !- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC !- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L07868; AAB59446.1; -.
 CC HSP; P11362; 1FGK.
 CC Genew; HGNC:3432; ERBB4.
 CC MIM; 600543; -.
 CC InterPro; IPR000494; EGF_L_domain.
 CC InterPro; IPR000719; Euk_pkinase.
 CC InterPro; IPR002174; Furin-like.
 CC InterPro; IPR001245; Tyr_pkinase.
 CC InterPro; IPR004019; YLP motif.
 CC Pfam; PF00069; pkinase; 1.
 CC Pfam; PF00757; Furin-like; 1.
 CC Pfam; PF01030; Recep_L_domain; 2.
 CC Pfam; PF02757; YLP; 2.
 CC ProDom; PD000001; Euk_pkinase; 1.
 CC SMART; SM00261; FU; 4.
 CC SMART; SM00219; Tyrc; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing.
 KW SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT PROTEIN KINASE.

NP BIND	724	732	ATP (BY SIMILARITY).
FT BINDING	751	751	ATP (BY SIMILARITY).
FT ACT_SITE	843	843	BY SIMILARITY.
FT DISULFID	189	197	BY SIMILARITY.
FT DISULFID	193	205	BY SIMILARITY.
FT DISULFID	213	221	BY SIMILARITY.
FT DISULFID	217	229	BY SIMILARITY.
FT DISULFID	230	238	BY SIMILARITY.
FT DISULFID	234	246	BY SIMILARITY.
FT DISULFID	249	258	BY SIMILARITY.
FT DISULFID	282	289	BY SIMILARITY.
FT DISULFID	293	304	BY SIMILARITY.
FT DISULFID	308	323	BY SIMILARITY.
FT DISULFID	326	330	BY SIMILARITY.
FT DISULFID	503	512	BY SIMILARITY.
FT DISULFID	507	520	BY SIMILARITY.
FT DISULFID	523	532	BY SIMILARITY.
FT DISULFID	536	552	BY SIMILARITY.
FT DISULFID	555	569	BY SIMILARITY.
FT DISULFID	559	577	BY SIMILARITY.
FT DISULFID	580	589	BY SIMILARITY.
FT DISULFID	593	614	BY SIMILARITY.
FT DISULFID	617	625	BY SIMILARITY.
FT DISULFID	621	633	BY SIMILARITY.
FT MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES	1294	1294	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD	138	138	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	174	174	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	181	181	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	253	253	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	358	358	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	410	410	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	473	473	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	495	495	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	548	548	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	576	576	N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD	620	620	N-LINKED (GLCNAC) (POTENTIAL).
FT VARSPLIC	626	648	NGPTSHDCIYPTGHTGHTSLPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B).
SEQ	SEQUENCE	1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;	

Query Match 43.4%; Score 2971.5; DB 1; Length 1308;
 Best Local Similarity 45.2%; Pred. No. 2,2e-152;
 Matches 612; Conservative 187; Mismatches 373; Indels 181; Gaps 30;

QY	9	WGLLLALLPPGAA-----STQVCTGTDMLKRLPASPETHLDMLRHLYQGQVVGQNLIELTY 64
Db	8	WVWSLLVAAGTVQPSDSQSVCAQTENKLSLSLDLEQQYRALRYKYYENCEVVMGNLEITS 67
QY	65	LPTNASLSFLQDIQEVQGVYLIHNOVRQVPLORLRIVRGTQLFEDNYALAVLDNGDPLN 124
Db	68	TEHNRDLSFLRSVREVTGVLVALNQFPLPLENRIIRGTGLYEDRYALALFLNRYKDG 127
QY	125	NTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDIFKNNQLALT1 184
Db	128	NF-----GLQELGLKXLTTEILGGVYDQNFCLCYADTIHWDIVRNWPSNLTIV 178
QY	185	DTNRSRACHPCSPMCKGRMGESSEDCOSLTRTVCAAGC-ARCKGPLPTDCCHEQCAAG 243
Db	179	STNGSGGGRCHKSCGTG-RCWGPTEHNCOTLRTVCAEQCDGRCYGPVYSDCHRECAAG 237
QY	244	CTGPKHSDCLACHENHSGICELHCP-ALVFNFTVSWLVRVPKVSASHLERYTGCASCV 302
Db	238	CSGPKDTCFACMFNFDSGACVTCQPTQFVYVNTTF-----QLEHNFNAKYTYGAFVCV 290
QY	303	TACPYNILSTDVSGCTLVCPHLNQVTAEDGTQRCCKSPKPCARVCYGLGMEHLREVRAV 362
Db	291	KKCPHFV-VDSSSCVACRACPSKMEV-BENGKMKCPCTDTCPKACDGTGTSLSMAQTV 348
QY	363	TSANIQEFAGCKKIFGSLAFIPESPDPGPNSTAPLOPEQLQVFTLEITGYLISAMP 422


```
Db 349 DSSNIDKFINCTKINGNLIFLVTHGIDPDYNAIEADPEKLNVRFTVREITGFLNIQSWP 408
Qy 423 DSLPDLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRLSRLRELSGLALHTHNLCLCF 482
Db 409 PNMTDFSVFSLVLTIGGRVLYSGLSLLILKQGITSLQFQSLKEISAGNIYITDNLCLY 468
Qy 483 VHTVPWDLFRNPQHALLHTANRPEDECVGEGLACHQICARGHCWGPPTQCWCSCFLR 542
Db 469 YHTINWTLFTSTINRIIVIRNKAENCTAEGMVCNHLCSGDCGPGPDQCLSCRRFSR 528
Qy 543 GOECVEECRVQLGLPREYVNAHCLPCHPEOP-ONGSVTCFPGPEADOCVACAHYKDDPF 601
Db 529 GRICIESCNLYDGEFREENGISICECDPQCEKMEDEGLLTCGPGPDNCTCKSHFKDGN 588
Qy 602 CVARCPGKVPDLSPMTWPKFPDEEGACOPCPINCTHSCVLDLDDKC-----PA 650
Db 589 CVEKCPDGLQANSF--IFKVADPDRECHPCNCTQCGNGPTSHDCIYVPMWTHSTLQP 646
Qy 651 EORASPLTSIVSAV-GILLVVVLGVFGILIKRQOKIRKYTWBRLLOETELVEPLTPS 709
Db 647 HAR-TPL--IAAGVJGGFLVILVGLTFVAVYRRKSIK-KKRALRRFL-ETELVEPLTPS 701
Qy 710 GAMPNQAQRLIKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 769
Db 702 GTAPNQAQRLIKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKA 761
Qy 770 NKEILDEAYVMAGVSPYVSRLLGLTSTVQLVTQLMYPYGCLLDHDVRENRLGSGDOLL 829
Db 762 NVEFMEALIMASDHDPHLVLLGVCLSTQTLVTLQMLPHGCLLVEYVHEHKDNIGSQLL 821
Qy 830 NWKQIAGKMSYLEVDRLVLRHDLAARNVLKSPNHVKITDGLARLLDIDETEHADGK 889
Db 822 NWKQIAGKMSYLEVDRLVLRHDLAARNVLKSPNHVKITDGLARLLDIDETEHADGK 881
Qy 890 VPIKWMALLESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKERLPQ 949
Db 882 MPKWMALLESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKERLPQ 941
Qy 950 PPICTIDVYIMVKCWMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEED-LGPAASPLDS 1008
Db 942 PPICTIDVYIMVKCWMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEED-LGPAASPLDS 1001
Qy 1009 TFYRSLLDDMDGLVDAAEYLVPOQGFCDPAPGACGMVHHRHSSTSGGGDLTLG 1068
Db 1002 KFFQNLDEEDLEMDMAEYLVLP-QAFNIPPP-----IYTSRAIDSNRS-----EIG 1049
Qy 1069 LEPSEEEAPRS-----PLAP-SEGAGSDVFDGDLGMG 1099
Db 1050 HSPPPATPMGNOFVYRDGGFAAEQGVSVYPRAPTSTIPEAPVAQGTATIFDDSCNG 1109
Qy 1100 AAKGLQLPHTDPSPQLQRYSDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVR 1152
Db 1110 TLRKPVAPHVQEDSSTORYSADPTVFAPERSPRGELDEEGYMTWRDKPKQYLNPNVE-- 1167
Qy 1153 PQPSPREGPLPAARPAGATLAKTLSPGKGVVGVKVFAGGAVENPEYLTPOGGAAPQ 1212
Db 1168 -----ENPVSRR-----KNGDLQ-----ALDNPETHNASNG----- 1194
Qy 1213 PHPPPA-----FSPAFDNLVYWDQDPPERGA 1238
Db 1195 --PKADEYVNEPLYLNTFANTLGAELKKNILSMPEKAKKAFDNDPDYWNHSLPPRST 1252
Qy 1239 --PSTFKGTPT-----AENPEYL 1255
Db 1253 LQHPDYLQEVSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ERR4 RAT
ID ERR4 RAT
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RL [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RL [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Fehner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF041838; AAD08899.1; -
EMBL; U52531; AAC53051.1; -
HSSP; P11362; 1FGK.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_Pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR004019; YLP motif.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; Pkinase; 1.
Pfam; PF01030; Recep_L_domain; 2.
Pfam; PF02757; YLP; 2.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_pkinase; 1.
```

DR SMART; SM00261; FU; 4.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_TOR; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 995 PROTEIN KINASE.
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 P.T -> S.T (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 43.2%; Score 2952; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 2 5e-151;
Matches 610; Conservative 194; Mismatches 382; Indels 168; Gaps 30;

QY 1 MELA-ALCRWGLLL--ALLPQGAASQVCTGDMKRLRLPASPETHLDMLRLHYLQGCQVQ 57
DB 1 MKLATGLWVGSLVAARTVQPSASQSVCAQTENKLSLSLEQQVRLRKYKYEVCVVM 60
QY 58 GNLEITYLPTNASLFLQDIQSVGVLTAAHNOVQVPLORLIRIVRGTLFEDNYALVL 117
DB 61 GNLEITSYHNRLSFLRSIREVTGYVLVALNQRYLPLENRIIRGTGKLYEDRYALAI 120
QY 118 DNGDPLNNTPTVTGASPGGLRELQLRLSLEILKGLVLIORNPOLCVQDTILWKDIFHKON 177
DB 121 LNYRKDGNF-----GLQELGLKNLLEILNGGVYVDONKFLCYADTIHQWDIRNPW 171

QY 178 QLALTLIDTNRSRACHPCSPMCKGSRGRCWGESSEDCQSLTRTVTCAGC-ARCKGPLPTDCC 236
DB 172 PSNMTLVSTIGSSGCGRCHKSCGTG-RCWGPTENHCOTLRTVTCABECDCRCYGPVYSDCC 230
QY 237 HQCAAGCTGPKHSCLACLAHFNHSGICELHCP-ALVFNFTVSVFLRVKPVKVSASHLEY 295
DB 231 HREACAGCGSGPKDTDCFACMFNDSGACVTCQCPQTFVYNPTTF-----QLEHNFNAKY 283
QY 296 TFGASCVTACPNYLSTVGSCVLCPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEH 355
DB 284 TYGAFCVKCKPHNFV-VDSGCCVRACPPSSKMEV-BENGIMCKPCPTDIPCACDGI GTGS 341
QY 356 LREVRAVTSANIQEFAGCKKIFGSLAFLPESPDGDPASNTAPLQPELOLVFETLEBITGY 415
DB 342 LMSAQTVDSNIDKFNCTCKINGNLIFLVTHGDPYNAIDAI DPEKLVNFRVREITGF 401
QY 416 LVISAWPSLPDLVSFQNLQVIRGRILNKGAVSLTQGLGIGLWGLRSRLRELGSGALAIH 475
DB 402 LNIQTWPPNMTDFSVFNSLVITIGRVLVSGLSLLILKQOGITSLQFSLKEISAGNIYIT 461
QY 476 HNTHLCFVHTVPNDOLFNPHQALLHTANRPEDECVGEGCLACHOLCARGHCWGPQTQCV 535
DB 462 DNSNLCCYHTINMTTLFSTVNQRIIVIRNRRRAENCTAEGMVCNHLCSNDCGCGPDPQCL 521
QY 536 NCSQFLRGQECVECEKRVLQGLPREYVNAHRLCPCHPECP-ONGSVTCFPGPADQCVACA 594
DB 522 SCRRFSRGKICIESCNLYDGEFREFENGSI CECDSQCEKMEDEGLLTCHGPGDNCCKS 581
QY 595 HYKDPFFCVARCPGKVPDLSYMPIWKFPDEEGACQPCPINTHSCVLDLDDKGC----- 648
DB 582 HFKDGPNCVEKCPVLQGANSP--IFKYADQRECHPCHPNCTQCGNCTSHDCIYYPWT 639
QY 649 -----PAEQRASPLTSIYSAVV-GILLVVVLGVGVGILLKRRQOKIRKYMRLLOTEL 702
DB 640 GHTSLPQHAR-TPL--IAAGVIGGLFVIMALTFAVYVRRKSIK-KKALRRFL-ETEL 694
QY 703 VEPLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVL 762
DB 695 VEPLTPSGTAPNQAQRLILKETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKILN 754
QY 763 ENTSKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTOLMPYGLCLDHLVRNRRGR 822
DB 755 ETTGPKANVEFMDEALINASVDHPLVRLGLVCLSPITQIQLVTLMPHGCLLYVHEHKN 814
QY 823 LGSQDLNKNQCIKAGMSYLEVDVRLVHRDLAARNVLKSPNKHVKTDFGLARLLDDETE 882
DB 815 IGSQDLNKNVCQIAKGMYLEERLVRDLAARNVLKSPNKHVKTDFGLARLLDEGEKE 874
QY 883 YHADGGKVPICKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLE 942
DB 875 YNADGGKVPICKMALECIHYRKFTHQSDVWSYGVTVWELMTFGKPYDGIPTREIPDLLE 934
QY 943 KGERLPQPICTIDVYIMVKWCMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEB-LG 1001
DB 935 KGERLPQPICTIDVYIMVKWCMIDSECRPRFRELVSFSEFMRARDPQRFVVIQNEB-LG 994
QY 1002 PASPLDSTFYSLLEDDMDGLVDAEYLVPOQGFCDP----- 1041
DB 995 LPSPNDSRFQNLLEDEEDLMDMAEYLVLP-QAENIPPIYTSRTRIDSNRSEIGHSP 1053
QY 1042 ---AFGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEAGSVDVFGDGLGM 1098
DB 1054 PAYTPMSGSQFVYQGGFATQGG---MPMPYATTTSTIPEAPVA--QGATAEMFDDSCCN 1108
QY 1099 GAAGLQSLPHDPSFLORYSEDPTVPLPS-----ETDGVVAPLTCSPQBEYVNPQDV 1151
DB 1109 GTLRKPVVPHVQEDSSTQRYSDPTVFAPERNPRAELDBEGYMTMHDKPKQEYLNVPV- 1167
QY 1152 RPQPSPREGLPAARAPAGATILERAKTLSPGKNGVVKDFAFGGGAVENTPEYLTPOGGAAP 1211
DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPYHSASSG----- 1194
QY 1212 QPHPPA-----FSPAFDNLVYWDQDPPERG 1237


```
Db 404 EITGVLMWNNPNNMTSLVFQNLLEIRGRTTFGRGFSFVVQVRHLQWLGLSLRSLKEVSA 463
Qy 470 GLALIHNTLHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVEGELACHQIACRGHCWP 529
Db 464 GNVILKNTLQRYANTINWRFLRSEDOSIEYDART-----ENQTCNNCESDGCGWGP 516
Qy 530 GPTQCVNCSQFLRGQECVEECRVLQGLPREVYNARHCLPCHPEQOPQNGSVTCFGEADQ 589
Db 517 GPTCVCSLHVRGRCVASCNLLQGEPREAQVDCRCVQCHQECVLQVDTSLTCYGPAN 576
Qy 590 CVACAHYKDPFCVACRSGVKPDLSPYPIKFPDEBAGACQPCINCTHSCVDLDDKCP 649
Db 577 CSKSAHFQDGPQICPRCPHGILGDGDTL-IWKYADKMGQCQPCQHCNCTQGCSSGPGLSGR 635
Qy 650 AEQRASPLTSSVAVGVLLVVLGWFGLIKRRQKIRKYTMRLQLQETELVEPLTPS 709
Db 636 GD-IVSHSSLAVGLVSGGLTIVVALLIVLURRRRIK-RKRTIRCLLQELVELVEPLTPS 693
Qy 710 GAMPNQAOMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKA 769
Db 694 GOAPNQAFLRIKTEFKKRVLSGAGFTVYKGLWNPFDGENIRIPVAIKVLRRENTSPKV 753
Qy 770 NKEILDEAYNAGVSPVSRLLGLICTSTVQLTQMLPYGCLLDHVRNRRGLSGDQL 829
Db 754 NOEVLDEAYNMAVDHPHVCRLGLICTSAVQLTQMLPYGCLLDYVVRQHOERICQWLL 813
Qy 830 NWCQIAGMSYLEDVRLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGK 889
Db 814 NWCQIAGMNYLEERHLVLRDLAARNVLLKNPNHVKITDFGLSKLTADKEVQADGK 873
Qy 890 VPIKWMALLESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 949
Db 874 VPIKWMALLESILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQ 933
Qy 950 PPICTIDVYMTVMKWMIDSCRRPRFRELVEFSFMSARDPQRFVVIQONEDLGPASPLDST 1009
Db 934 PPICTIEVYMTILKWMIDSPSRPRFRELVEFSQMARDPQRYLVIQO---NLPFSLSDRR 990
Qy 1010 FYRSLLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGL 1069
Db 991 LFSRLSSDD--DVVDADEYLLPYKRI-----NRQGS----- 1020
Qy 1070 EPSEEEAPRSLAPSEAGSDVFDGLGMAKGLQSLPTHDPSPQLQRYSEDPV-PLPS 1128
Db 1021 -----EPCIPPTGCH-----PVRENSITLRNISDPTQNALEK 1051
Qy 1129 ETDGYVAPLTCSPQPEYVQPDVRPQ-----PSPRE-----GPLP-AARPAGATLER 1175
Db 1052 DLDDG-----EYVNPQGETSSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFSR 1101
Qy 1176 AKTLPSPGKGVVQVFAFGGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPE 1235
Db 1102 PEYLTNQNLSU---PLVSSGSMDDPDY---QAG-----YQAAF-----LPQ 1136
Qy 1236 RGAPPSTFKGTPTAENPEYLG 1256
Db 1137 TGALTNGMFLPAAENLEYLG 1157
```

RESULT 9

```
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
GN ERBB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoch M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -|- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -|- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -|- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -|- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -|- SH2: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -|- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1
DR Pfam; PF00757; Furin-like; 1
DR Pfam; PF01030; Recept_L_domain; 2
DR ProDom; PD000001; Euk_pkinase; 1
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW
```

[illegible]

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbb-3 precursor (EC 2.7.1.112)
 DE (c-erbB3)
 GN ERBB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85: 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fehner P.W., Kim S.S., Corbett J.A.;
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DOMAIN: THE CYTOSOLIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U29339; AAC28498.2; -
 DR EMBL; U52530; AAC53050.1; -
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk pkinase; 1.
 DR SMART; SM00261; FU_5
 DR SMART; SM00219; TyrK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT BIND 721 740
 FT BIND 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 556 573
 FT DISULFID 576 585
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 FT CONFLICT L -> P (IN REF. 31).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AASF2402BBFDF1E CRC64;

Query Match 34.1%; Score 2329.5; DB 1; Length 1339;
 Best Local Similarity 40.5%; Pred. No. 7.3e-118;
 Matches 523; Conservative 173; Mismatches 431; Indels 163; Gaps 36;
 Qy 3 LAALRWGLLLALLPPGAA---STQVCTGTDMLRLPASPETHLDMLRHLYGCGQVOGN 59
 Db 7 LQVLC-----FLSLARGSEMGNQAVCPGTLNGLSVTGDADNQYQTLKLYEKCEVVMGN 62
 Qy 60 LELTYLPTNASLSFLQDIOEQGVYVLIAHNOVROVPLQRLRVRGTLQDFEDNYALAVLDN 119
 Db 63 LEIVLTGHNADLSFLQVIREVTGYVLVANNEFSLPLPNLRVVRGTQVYDGKFAIFVM-- 120
 Qy 120 GDPLNNTTPVTGASPGGLURELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHNKOL 179
 Db 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR--- 170
 Qy 180 ALTLIDTNRSPACHPCSPMKSGRCWGESSEDCQSLTFTVCAGGC-ARCKGPLPTDCCHE 238
 Db 171 GAEIVVXNGANGCPPCEHEVCCKG-RCWGGPDDCQLTKTKTICAPQCNGRCFCGPNQCCHD 229
 Qy 239 QCAAAGCTGPKHSDCLACLHFNHSGICELHCP-ALVFNNTVSVFWRVPRKVSASHLERYTF 297
 Db 230 ECAGCGSPQDTDCFACRRFNDSGACVRCRCEPLVYNKLTFT-----QLENPHTKYQY 282
 Qy 298 GASCVTACPNYSLTDVSGCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLQMEHLR 357
 Db 283 GGVCVASCPHNFV-VDQTFVCVRCACPPDKMEVD-KHGLKNCEPCGGCLCPKACGTCGSG--S 338
 Qy 358 EVRAVTSANIOEFAGCKKIFGSLAFLPESFDCDPRASNTAPLOPEQLQVETLEEITGYLY 417
 Db 339 RYQTVDSNIDGTFVNCCKILGNLDFLITGLNVDPMHKIPALDPEKLNFRVREITGYLN 398
 Qy 418 ISAWPDSLPDLVSFQNLQVIRGRIHLNGAYS-LTLQGLGISWLGURSLRSLGSLALIH 476

Db 399 IQSWPPHNFVSFNLTTIGRSLYNRFGLLMKLNVTSLGFRSLKEISAGRVYISA 458
Qy 477 NTHLCFVHTVPDQFLRNPHOALLHTA--NRPEDECVGEGGLACHQLCARGHCGPGPTQCV 535
Db 459 NOQLCYHSLNWKTRLLRPSSEERLDIKYDRPLGECLAEGKVDPLCSSGGCGPGPGQCL 518
Qy 536 NCSOFLROGECVEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFPGPADQVACAH 595
Db 519 SCRNYSREGVCVTHCNFLQGPREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAH 578
Qy 596 YKDPPEFCVARGPSGVKPDLSYMPKTFPDEGACQPCPINTHSC--VLDLDDKGCPAEQR 653
Db 579 FRDGHVCNCPHGILG--AGPIYKIPDAQNECRPCHENCITQCGNPELQDCLGQAQAEVL 636
Qy 654 ASPLTSIVSAVGVLLVVLGVVFGILIKRRQOKIR-KYTMRRLLQETELVEPLTPSGAM 712
Db 637 MSKPHLVIAVTVG--LAVILMILGSGFLYWRGRRIQKRAMRRLYERGESIEPLDPS--EK 693
Qy 713 PNQAMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKE 772
Db 694 ANKVLARIFKETELRKVLKVLGSGVFGTVHKGWIPEGESIKIPVCIKVIEDKSGRSQFA 753
Qy 773 ILDEAYVMAGVSPVSRLLGICLTSTVQLVTLQMLPYCCLLDHVRNRRGLSGQDLINWC 832
Db 754 VTDHMLAVGSLDHAHIVRLGLLCPGSSQLVTVQLPLGSLLDHVKQHRRETLGQQLLNWG 813
Qy 833 MQIAKMSYLEDLVLRHDLAARNVLKSPNHRKITDGLARLLDIDITEYHADGKVPKI 892
Db 814 VOIAKGMYLEHSMVHRDLALRNVLKSPSOVQVADFGVADLLPDDKQLLHSEAKTPI 873
Qy 893 KMALESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI 952
Db 874 KMALESIFHGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAETPDLLEKGERLAQPI 933
Qy 953 CTIDVYMWKCMIDSECRPRFRELSEFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYR 1012
Db 934 CTIDVYMWKCMIDENIRTFKELANEFTFMRDPRYLVIKRA--GPCTP--PAAEP 990
Qy 1013 SLEDDDMGDLVDAEYLVPOQGFCDPAPGAGWVHRRSSRSGGDLTLGLEPS 1072
Db 991 SVLTTKEL-----QEALEPEL-----DLDDLEAE 1016
Qy 1073 EE-----EAPRPLAPSEG-----AGSDVFDGLGMAKAGL 1104
Db 1017 EBLGATSLGALSILPTGTLTPRGSQSLSPSSGYMPMNOSSGEACLDASVILGREGQFS 1076
Qy 1105 QSLPHTDPSPLQRYSEDPVLPSETQYV-----APL-----TC-----SPOPE----- 1144
Db 1077 RPTSLH-PIPRGR-----PASESSEGHVTGSAELQEKVSVCRSRSRSPRPGDSA 1128
Qy 1145 YVNPQDVRPOPPSPREGP-----LPAARPAGATLERAKTLP--GKNGVY----- 1187
Db 1129 YHSQRHLLTPVTLPSPLGLEEDCGVMPDTHLRCGASSREGTSSVGLSSVLGTTEE 1188
Qy 1188 -KDVFAFGGAVENPEYLTPOGGAAPQHPH 1216
Db 1189 DED-----EYEYNNRKRGRSP-PRPP 1209

RESULT 11

EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB)
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RN [1] NCBI TaxID=7227;
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RL subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RL hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RL Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J.J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
[7]
RN
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong,
RX MEDLINE=85137936; PubMed=2983322;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A *Drosophila* genomic sequence with homology to human epidermal
RT growth factor receptor.";
RL Nature 314:178-180(1985).
[8]
RN
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
[9]
RN
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perlmutter N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
RT EGF receptor.";
RL Cell 89:13-16(1997).
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN LARVAE, IN LARVAE,
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, ANLAGEN OF
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK IN ADULTS. EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF052754; AAC08536.1; -;
CC EMBL; AF052753; AAC08536.1; JOINED.
CC EMBL; AF052754; AAC08535.1; -;
CC EMBL; AF052752; AAC08535.1; JOINED.
CC EMBL; K03054; AAA51462.1; -;
CC EMBL; K03417; AAA51460.1; -;
CC EMBL; K03416; AAA50965.1; -;
CC EMBL; K03418; AAA51461.1; -;
CC EMBL; AF109077; AAD26134.1; -;
CC EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; -;
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; -;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -;
DR EMBL; AE003454; AAF46732.1; -;
DR EMBL; X02293; CAA26157.1; -;
DR EMBL; X78920; CAA55523.1; -;
DR EMBL; X78918; CAA55521.1; -;
DR EMBL; X78919; CAA55522.1; -;
DR PIR; A00640; GORPE.
DR HSSP; P11362; 1RGK.
DR Flybase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 28.1%; Score 1922; DB 1; Length 1426;
Best Local Similarity 32.5%; Pred. No. 6.2e-96;
Matches 465; Conservative 185; Mismatches 436; Indels 344; Gaps 42;

QY 24 QVCTGTDMKRLPASPETHDMLRHLYGQCQVVOGNLELTLYLPT-NASLSFLQDIQVOG 82
DB 100 KICIGTGRSLSPVSPNKEHHYHNLDRYNTCTYVDGNLKLTLWLPNENLDSFLDNIREVTG 159
QY 83 YVLIANQVQVPLQRLRIVRGTLQF-----EDNYALAVLDNGDPLNNTTPVTGASPGSL 137
DB 160 YILISHVDVKKVFPKLIIRGRTLFLSVEEKYALFV-----TYSQM 203
QY 138 RELQLRSITELTKGVLIQRPOLCYQDTILWKDIFHQNLALTLIDNRSRACHPCSP 197
DB 204 YTLIPLDLRVLDVNGVGFPHNNYNLCHMRTIQMSEIVSNGTDAYVNYDFATPERECPKCHE 263
QY 198 MCKGSRWCSESDCQSLTRTVACAGGCA--RCKGPLPTDCCHCEQAAGTCGPKHSDCLAC 255
DB 264 SCTHG-CWGEQPKQCFKSLTCSQACAGGRCYGPKECHLFCAGGCTGTQKDCIAC 322
QY 256 LHFNHSGICELHCPALVFNFTVFWLRVPKVSASHLERYTFGASCVTACPVNYLSTDVG 315
DB 323 KNFFDEAVSKEECPPMRYNPTTVVLETNPE-----GKYAYGATCVKECP-GHLRDRNG 375
QY 316 SCTLVCPVLRHNEVTAEDTQRCCKSPCARVCVGLGMEHRLREVRVAVTSANIQEPAGCKK 375
DB 376 ACVRSCPDQKMDKGE-----CVPCNGPCPKTCGVTVLH-----AGNIDSPNCTV 422
QY 376 IFGSLAFLPESFDG--DPASNTA-----PLQPEQLQVFTELEITGYLISAWPDSLDP 427

```
Db 423 IDGIRILQDFSGFQDYVANYTMGPRIYIPDPERREVFSTVKEITGYLNIEGTHPQPRN 482
Qy 428 LSVFQNLQVIRGLHNGAY-SLTQGLGIGSWGLRLSRLGSLALIHNNHLCFVHTV 486
Db 483 LSYFRLNETHGRQLMESFAALAIVKSSLYSLEMRNLKQISSGSGVVIQHRDLGYVSI 542
Qy 487 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGCHGWPGPTQCVNCSQFLRGQEC 546
Db 543 RWPALQKEPEQKVMNENLRADLCEKNGTI CSDQCNEDCGWAGDQCLTKCNFNFGTC 602
Qy 547 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 606
Db 603 IADCGYISNAYK--FDNPTKICHPECR-----TCNAGADHQCQCVHVRDQGHVSEC 654
Qy 607 P-----SGVK-----PDL 614
Db 655 PKNYNDRGVCRECHATCDGCTGPKDTTIGACTTCNLAIINNDATVRCLLKDDKCPD- 713
Qy 615 SYMPIWKF--PDEGACOP-----CPI-----NCTH----- 638
Db 714 GY---FWEYVHPQEGSLPLAGRAVCRKCHPLCELCTNYGHEOVCSKCTHYKREOCET 771
Qy 639 -----SC-----VDLDDKG----- 647
Db 772 ECPADHYTDEQRECFQHPPECNGCTGPGADDCKSCRNFKLFDANETGPNYSTWFNCTS 831
Qy 648 -CPAEQR-----ASPLTS-----IVSAVVGILLVVLGVVFGI 679
Db 832 KCPLMRHVNYYQTAIGPYCAASPRSSKITANLDVNMIFITGAVLPTICILCV--T 889
Qy 680 LIKROOKIRKYT--MRLLOTELVELTPTSGAMPNQAOHRIKELTELKRVKVLGSGAF 737
Db 890 YICQKQKAKKETVMYTMALSGCEDSEPLRPSNTGANLCKLRIIVKDAELRKGVLGMAF 949
Qy 738 GTVYKGTWIPDGENVKIPVAIKRENTSPKANKEILDEAVYVAGVSGPYVSRLLICLT 797
Db 950 GRVYGVVPGENVKIPVAIKELKSTGAESSEFEAREAYIMASEEHVNLKLLAVCMS 1009
Qy 798 STVOLVTQMPYGCLLDHVRENRRGLSGQDILLNMQITAKMSYLEDLVRLVHRLAARNV 857
Db 1010 SOMMLITQLMPLGLCLLDVVRNRRDKIGSKALLNWSQIAKMSYLEEKRLVHRDLAARNV 1069
Qy 858 LVKSPNHVKITDFGLARLIDIDETEHADGKQVPIKMALESILRRRPTHOSDWSVGV 917
Db 1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAAGKMPKIMALECIRNRVFTSKSDVWAFGV 1129
Qy 918 VWELMTGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVTYIMVMVKMIDSECRPRFE 977
Db 1130 IWELMTFGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAAVRPTKQ 1189
Qy 978 LVSEFSNARDPQRFVQVNEIDLQ--PASPLDSTFYRSLLEDD--DMGDLVDAEYLVP 1032
Db 1190 LTTVFASFARDPGRYLAIPGDKFRLPA-----YTSQDEKDLIRKLAPTPTDGSALAK 1242
Qy 1033 QOQFFCPDPARGAGVMVHRRSSRSTSGGDLTLGLEPSEEA-----RSLPASEG 1086
Db 1243 PDDYLQKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSS 1281
Qy 1087 AGSDVDFG--DLGMAAGKGLQSLTHDPSPQRYSEDPTVPLPSETGDYVAPLTCSPQ 1143
Db 1282 TGDDESSAREVGVNLR-----LDLPVEDDYLMP--TCQGP 1319
Qy 1144 EYVNPQDVRPOPSPREGPIPAARPAAGATLERAKTLSPGKNGVVKQVFAFGAVENPEYL 1203
Db 1320 NNNNNNN-----NPNQNNMAAVGAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1204 ----TPGGGAAPQH-----PPPAFSP-AFNLUYWD 1230
Db 1359 LNAQTLGVGSSPTQTGTIGIPVMGGPGTMEVKVPMGSEPTSSDHEHYND 1408
```

RESULT 12

```
ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFLV.
CC HSP; P11362; IFKG.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS00111; PROTEIN KINASE DOM; 1.
CC Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT_SITE 257 257 BY SIMILARITY.
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
CC
CC Query Match 25.6%; Score 1749.5; DB 1; Length 634;
CC Best Local Similarity 52.3%; Pred. No. 4.5e-87;
CC Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
```

Qy 593 CAHYKDPFPCVACPSGVKFDLSYMPIWTKVPPDEGACQPCPINCTHSCVDLDDGCPAEQ 652

Db 3 CAHFIDGPHCVKACPCAGVGLGENDTL-VMKYADANAVCOLCHPNCTRGKGLGECP--- 58

Qy 653 RASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKVTMRRLLOTELVELPTPSGA 711

Db 59 NGSKTSPSIAGVVGGLLCLVVGIGLIGLYLRRR-HIVKRTLRLQLERLEVEPLTPSGE 117

Qy 712 MPNOAQMRILKETELRKVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771

```
118 APNOAHLRLKETETPKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
QY 772 EILDEAYMAGVSGPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENRLGSDLLNW 831
Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVLTQMPYGCCLLDYIREHKDNTGSOYLLNW 237
QY 832 CWOIAGMSYLEDVRLVRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPORVITDFGLAKLCADEKEHAGGGKVP 297
QY 892 IKWMALESILRRRFTHQSDVMSYGVTVMLMTFGAKPYDGPAREIPDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVMLMTFGSKPYDGPASBISSVLEKGERLPQPP 357
QY 952 ICTIDVYIMVKCMWIDSECRPRFRELSESRMARDQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKMARDPRLVLIQDGERMHLPSPDTSKF 417
QY 1011 YRSLLDEDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPSPL-----APSEGAGSDVFDGLGMGAAGLQSLPHTDPPSPQRYSEBDPTVP 1125
Db 450 -----SRTPLLSLSATSNNSATNCID-----RNQCGHPVREDSPVQRYSDPTCN 495
QY 1126 LPSET--DGYVAPLTCSPQEVVQPDVRPQPPSPREGPLPAARAGATLERAKTLPCK 1183
Db 496 FLEESIDGFL-----PAPEYVQ--LMPKKPS-----TAMVQ 526
QY 1184 NGVVKDF-----AFGGAIVENPEYLTPOGGAAPQHPPPAFSPAFDNLXY 1228
Db 527 NQIYNNISITAIKSLPMDSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVPFESSPY 578
QY 1229 WDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLOVP 1260
Db 579 WIQSGNHQINLNDPDYQDFLPNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=H;
RC MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserre G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
```

```
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A06644; TVYUH.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
FT SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;
Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.4e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 593 CAHYKDPFPFCVAPCSGVKPDLSYNPIWKFPDEACQPCPCINCHTSCVDLDDKGCPAEQ 652
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVCQLCHPNCCTRGCKGPGLEGCP--- 58
QY 653 RASPLTSIVSAVV-GILLVWVGVVGLIKRQOKIRKYTWRRLLQETELVEPLTPSGA 711
Db 59 NSGKTPTSIAAGVVGGLCLLVVGLIGLYLRRR-HIVKRTLRLLQERLVEPLTPSGE 117
QY 712 MPNOQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKELRENTSPKANK 771
Db 118 APNOAHLRLKETETPKVKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
QY 772 EILDEAYMAGVSGPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENRLGSDLLNW 831
Db 178 EILDEAYMASVDNPHVCRLLGICLTSTVLTQMPYGCCLLDYIREHKDNTGSOYLLNW 237
QY 832 CWOIAGMSYLEDVRLVRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 891
Db 238 CVOIAGMNYLEERLVRDLAARNVLKTPORVITDFGLAKLCADEKEHAGGGKVP 297
QY 892 IKWMALESILRRRFTHQSDVMSYGVTVMLMTFGAKPYDGPAREIPDLEKGERLPQPP 951
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVMLMTFGSKPYDGPASBISSVLEKGERLPQPP 357
QY 952 ICTIDVYIMVKCMWIDSECRPRFRELSESRMARDQRFVVIQ-NEDLGPASPLDSTF 1010
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKMARDPRLVLIQDGERMHLPSPDTSKF 417
QY 1011 YRSLLDEDDMGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRHRSSTRSGGDLTLGLE 1070
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEEAAPSPL-----APSEGAGSDVFDGLGMGAAGLQSLPHTDPPSPQRYSEBDPTVP 1125
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 Seconds
(without alignments)

5347.444 Million cell updates/sec

Title: SEQ4-250-264-14

Perfect score: 6839

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGIDVVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6191	90.5	1259	6	O18735
2	3119	45.6	1209	11	Q9GX70
3	3090	45.2	1210	11	Q9EP98
4	2724	39.8	1165	13	Q9YH40
5	2704.5	39.5	1137	13	Q9W6F6
6	2284	33.4	1328	13	P79754
7	2011.5	29.4	1433	5	Q9B1H9
8	1770.5	25.9	419	4	Q9UK79
9	1739	25.4	367	11	Q8R2X1
10	1720	25.1	729	15	Q86712
11	1718	25.1	567	15	Q86714
12	1697.5	24.8	412	4	Q8WYV0
13	1653.5	24.2	962	15	O64895
14	1645	24.1	545	15	O85468
15	1481.5	21.7	655	11	Q9WVF5
16	1465.5	21.4	643	11	Q9ERV6

Q9Y1X8 ephydatia f
Q23821 caenorhabdi
Q26566 schistosoma
Q90836 gallus gall
Q9880 rattus norv
Q9PSH2 gallus gall
Q14256 homo sapien
Q923V5 rattus norv
P11776 xiphophorus
Q882W1 drosophila
Q93162 xiphophorus
Q9PVZ4 xenopus lae
Q9NJV5 biophalari
Q8UW85 paralichthy
Q93457 scophthalmu
Q9BUD7 homo sapien
Q8UW86 paralichthy
Q73798 xenopus lae
Q9U5A8 bombyx mori
Q8UW84 paralichthy
Q9YGH8 scophthalmu
Q8UW83 paralichthy
Q9QVW4 rattus sp.
Q9BG66 oryctolagus
Q9VD94 drosophila
Q91YM0 mus musculu
Q07912 homo sapien
O54967 mus musculu
Q9UMQ4 homo sapien

ALIGNMENTS

RESULT 1

O18735
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ErB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_taxID=9615;
RN [1]_taxID=9615;
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erB-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBAJ databases.
DR EMBL; AB008451; BAA23127.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match		90.5%;	Score 6191;	DB 6;	Length 1259;
Best Local Similarity		90.6%;	Pred. No. 0;		
Matches 1147;		Conservative 41;	Mismatches 66;	Indels 12;	Gaps 3;
QY	1	MELAAALRCWGLLLALLP	PGAASTQVCTGDMKLR	PASPETHLDMRLRHLY	YQCGVQVQGNL 60
DB	1	MELAAALRCWGLLLALLP	PSGAAGTQVCTGDMKLR	PASPETHLDMRLRHLY	YQCGVQVQGNL 60
QY	61	ELTYLPTNASLFLQDI	QIEVQGVLI	TAHNOVQLR	LRIVRGTQFDPEDNALAVLNG 120
DB	61	ELTYLPTNASLFLQDI	QIEVQGVLI	TAHNOVQLR	LRIVRGTQFDPEDNALAVLNG 120
QY	121	DPUNNTPTVTGAS	PGGLRELQRLSLTE	ILKGGVLIQRNPOLCY	QDQDTILWKDIFHKNNQLA 180
DB	121	DPLEGGIPAPGAAQGL	RELQRLSLTE	ILKGGVLIQRNPOLCHQD	TILWKDIFHKNNQLA 180
QY	181	LTLIDTNRSRACHPC	SPCKGRCWGESSE	QSLTRTYCAGCARCKG	PLPTDCHEQC 240
DB	181	LTLIDTNRSAKPPC	SPACKDAHCWASSG	DCQSLTRTYCAGCARCKG	PQPTDCHEQC 240
QY	241	AAGCTGPKHSDCLAC	LHFNHSGICELHCP	ALVFNFTVSFWLR	VPKVSASHLERYTFGAS 300
DB	241	AAGCTGPKHSDCLAC	LHFNHSGICELHCP	ALVYNTDTFESMPNFB	-----GRYTFGAS 294
QY	301	CVTACPYNYLSTDV	SGCTLVCPPLHNQEV	TAEDGTQRCCKSPCAR	VCYVYGLGMEHLREYR 360
DB	295	CVTSCPYNYLSTDV	SGCTLVCPPLHNQEV	TAEDGTQRCCKSPCAR	VCYVYGLGMEHLREYR 354
QY	361	AVTSANIQEFACKKI	FGSLAFELPES	FDGDPASNTAPLO	PEOLOVPETLEEITGYLYISA 420
DB	355	AVTSANIQEFACKKI	FGSLAFELPES	FDGDPASNTAPLO	PEQLRFEALEEITGYLYISA 414
QY	421	WPDLSLPDLSVFQ	NLOVIRGRI	LHNHNGAYSLT	LOGLGISWLGSLRLSGLALIHNTHL 480
DB	415	WPDLSLPNLSVFQ	NLRVIRGRLV	HDGAYSLT	LOGLGISWLGSLRLSGLALIHNRAL 474
QY	481	CFVHTVPMDQLFR	NPHQALLHTANR	PEDECVGEGACHOL	CARGHGWGPGTQCVCNCSQF 540
DB	475	CFVHTVPMDQLFR	NPHQALLHSANR	PEEBCVGEGLACP	-CAHGHGWGPGTQCVCNCSQF 533
QY	541	LRGQCEVRECVLQ	GLPREYVVARHCL	PCHPECOFONGSVT	CFGPEADOCVACAHYKDDP 600
DB	534	LRGQCEVRECVLQ	GLPREYVVDKRYC	LPCHSCQCPQNGSVT	CFGSEADOCVACAHYKDDP 593
QY	601	FCVARCPSGVKPD	LSYMPYWKFP	DBEGACQPCP	INCTHSCVDLDDKGCPCAEORASPLTSI 660
DB	594	FCVARCPSGVKPD	LSFMPYWKFADE	BEGTCQPCP	INCTHSCADLDEKGCPCAEORASPVTSI 653
QY	661	VSAAVGIILLVVLG	VVVGILIKRQOK	IRKRYTMRLLQ	ETELVEPLTPSGAMPNQAQMRI 720
DB	654	IAAVVGILLVAVVG	LVGLGILIKRRRQ	KIRKRYTMRLLQ	ETELVEPLTPSGAMPNQAQMRI 713
QY	721	LKETELRKVKVLG	SAGFVYVYKGI	WIPDGENVKI	PVAIKVLRNTSPKANKELIDEAYVM 780
DB	714	LKETELRKVKVLG	SAGFVYVYKGI	WIPDGENVKI	PVAIKVLRNTSPKANKELIDEAYVM 773
QY	781	AGVGSPPYVRLG	ICLTSTVQLVTQ	LMPYGCILLD	HVRENRLGSGODLLNMCQAKGMS 840
DB	774	AGVGSPPYVRLG	ICLTSTVQLVTQ	LMPYGCILLD	HVREHRLGSGODLLNMCVQAKGMS 833
QY	841	YLEDVRLVHRDLA	ARNVLKSNHVKI	ITDFGLARLLD	IDETEHADGGKVPDKWMALES 900
DB	834	YLEDVRLVHRDLA	ARNVLKSNHVKI	ITDFGLARLLD	IDETEHADGGKVPDKWMALES 893
QY	901	LRRRTTHOSD	WMSYGVYTWELMT	FGAKPYDGI	PAIREIPDLLEKGERLPQPPCTTDVYMI 960
DB	894	PPRRTTHOSD	WMSYGVYTWELMT	FGAKPYDGI	PAIREIPDLLEKGERLPQPPCTTDVYMI 953
QY	961	MVKCMWIDSECR	PRELVESEFS	RMARDPQREW	IQNEDLGPASPLDSTFYRSLLDDDM 1020
DB	954	MVKCMWIDSECR	PRELVAEFS	RMARDPQREW	IQNEDLGPASPLDSTFYRSLLDDDM 1013
QY	1021	GDLVDAEYLV	VPQGGFFCPD	PAPAGAGGV	MVHHRSSSTRSGGDLTLGLEPSEEPKSP 1080

Db	1014	GDLVDAEYLV	VPQGGFFCPD	PTPGAGGT	AHRRSSSTRNGGELTLGLEPSEEPKSP	1073
Qy	1081	LAPSEGAGSDV	PDGDLGMAAGK	GLSLPTHDSPLQ	RYSEDPTVPLPSETDGVVAPLTCS	1140
Db	1074	LAPSEGAGSDV	PDGDLGMAAGK	GLSLPTHDSPLQ	RYSEDPTVPLPSETDGVVAPLTCS	1133
Qy	1141	POPEYVNPQDVR	PQPPSPREGPL	PAARPAGATLER	-----AKTLPKNGKGVKDVFAFG	1195
Db	1134	POPEYVNPQDVR	PQPPSPREGPL	PAARPAGATLER	PKTLPKNGKGVKDVFAFGS	1193
Qy	1196	AVENPEYLT	PQGGAAAPQPHPP	PAFSPAFDNL	LYWDQDPPERGAPSTFKGTPTAENPEYL	1255
Db	1194	AVENPEYLA	PRGAAAPQPHPP	PAFSPAFDNL	LYWDQDPPERGAPSTFKGTPTAENPEYL	1253
Qy	1256	GLDVPV	1261			
Db	1254	GLDVPV	1259			
RESULT 2						
Q9QX70 PRELIMINARY; PRT: 1209 AA.						
ID	Q9QX70					
AC	Q9QX70;					
DT	01-MAY-2000	(T-EMBLrel. 13, Created)				
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)				
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)				
DE	Epidermal growth factor receptor.					
GN	EGFR.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.					
OX	NCBI_TaxID=10116;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RX	MEDLINE=90258888; PubMed=2342466;					
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,					
RA	Earp H.S.;					
RT	"A truncated, secreted form of the epidermal growth factor receptor is					
RT	encoded by an alternatively spliced transcript in normal rat tissue.";					
RL	Mol. Cell. Biol. 10:2973-2982(1990).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Petch L.A.;					
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Guttridge K., Dawson T.L., Earp H.S.;					
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; M37394, AAF14008.1; -					
DR	HSP; P11362; IFGK.					
DR	InterPro; IPR000494; EGFR_L domain.					
DR	InterPro; IPR000719; Euk_pkinase.					
DR	InterPro; IPR002174; Furin-like.					
DR	InterPro; IPR001245; Tyr_pkinase.					
DR	Pfam; PF00757; Furin-like; 1.					
DR	Pfam; PF00069; pkinase; 1.					
DR	Pfam; PF01030; Recep_L domain; 2.					
DR	PRINTS; PR00109; TYRKINASE.					
DR	ProDom; PD000001; Euk_pkinase; 1.					
DR	SMART; SM00261; FU; 3.					
DR	SMART; SM00219; TyrKc; 1.					
DR	PROSITE; PS00107; PROTEIN KINASE_ATP; 1.					
DR	PROSITE; PS00011; PROTEIN KINASE_DOM; 1.					
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.					
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.					
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEETFeCCIB7773 CRC64;					

Query Match 45.6%; Score 3119; DB 11; Length 1209;
Best Local Similarity 49.6%; Pred. No. 7.8e-227;

DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase.	
SQ	SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;	
	Query Match 45.2%; Score 3090; DB 11; Length 1210;	
	Best Local Similarity 49.1%; Pred. No. 1.2e-224;	
	Matches 628; Conservative 170; Mismatches 364; Indels 116; Gaps 24;	
QY	11 LLALLPPGAA--STOVCTGTDKMLRLPASPTHLDMLRLHLYGCGQVQGNLELTLYPTN 68	
DB	14 LLTALCAAGALEEKVKVCGTSGNLTQLGTTFEDHFLSLQRMVNNCEVNLGNLEITVYQRN 73	
QY	69 ASISFLQDLOEVQGVYLAHNVQVPLQRLIRVGTQTFEDNYALAVLDNGDPLNNTTP 128	
DB	74 YDLSFLTKTQEVAGYVLIALNTVERIPLENLQIRGNALYENTYALAILSN----- 124	
QY	129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYODTILWKDI-----FKNNQLALTLI 184	
DB	125 -YGTNRTGLRELPMRLNLEILIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMDL--- 180	
QY	185 DTRNSRACHPCSPMKGKSCWGESSEDCQSLRTVTCAGGCA-RCKGPLPTDCHEQCAAG 243	
DB	181 -QSHPSCKPCDPSCPNCGSCWGGGEENCKLTKIICAQOCSHRCGRSPSDCCHNQCAAG 239	
QY	244 CTGPKHSDCLACLHFNHSGICELHCPALVFNFTVFWLRVPKVVSASHLERYTFGASCVT 303	
DB	240 CTGPRESDCLVQKFODEATCKDTCPPLMLNPTTY-----QMDVNPPEGKYSFGATCVK 293	
QY	304 ACPPNYLSDVGSCTLVCLPHNQVEVTAEDGTORCEKSCPKARVCVYGLGWEHLREVRVAT 363	
DB	294 KCPRNVTVDHGSVCRAACGPDYEV-EEDGIRKCKDCGPKVCNGIGIGEFKDTLSIN 352	
QY	364 SANIQEFACKKIFGSLAFPLPSFDGDPASNTAPLOEQLQVPETLEEITGYLYISAWPD 423	
DB	353 ATNIRKFKYCTAISGDHLPLVAFKGDSTFTPLDPRELEILTKVKEITGFLLIQAWPD 412	
QY	424 SLPDLSVFONLOVIRGRIHLNAGYSLTLOGLGSWGLSLRSLGSLALIHNTLHLCFV 483	
DB	413 NWTDLHAFENLEIRGTQKQGFSLVAVGLNITSGLRSLKEISDGVIIISGNRLCYA 472	
QY	484 HTVPMQDLFRNPQHALLHTANRPEDECVGEGGLACHOLCARHGWGPGTQCVCNCSQFLRG 543	
DB	473 NTINMKLFGTNPQTKIWNNAEKDKCAVNHVCNPLCSSEGCWGPEDCVCSQNVSRG 532	
QY	544 QCVVEECRVLOGLPREYVNAHCLPCHPECCQONGSVTCFGEADOCVACAHYKDDPPFCV 603	
DB	533 RECVEKCNILEGEPREFVENSEICIQHPECLPQAMNITCTGRGPDNCIOCAHYIDGPHCV 592	
QY	604 ARCPGSKPDLVMPITWKFPEDEGACQPCPINTHSCVDLDDKGCPEAQRASPLTSIVSA 663	
DB	593 KTCPAGIMCENTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCVEWSPGKPIPSIATG 651	
QY	664 VVGILLVVLGVVFGI-LIKRQKQKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLK 722	
DB	652 IVGGLLFIW-VALGIGLFMRRRHVRKTRLLQRELRELVEPLTPSGEAPNQAHRLK 710	
QY	723 ETELKRVKVLGSGAGTGVYKGIWIPGENVKIPVAIKVRENTSPKANKEILDEAYVMAG 782	
DB	711 ETEFKKIKVLGSGAGTGVYKGLWIPEGEVKIPVAIKELREATSPKANKEILDEAYVMAS 770	
QY	783 VGSPPVSRLLGLCLTSVTQVLTQMLPYGCLLDHVRNRRGLSQDILLNWCMIKAGMSYL 842	
DB	771 VDNPHVCRLLGLCLTSVTQVLTQMLPYGCLLDVYREHKONIGSYQLLNCWCQVIAKGMNVL 830	
QY	843 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGKVPKIKMALESILR 902	
DB	831 EDRLVHRDLAARNVLKTPQHKITDFGLAKLGAEEKEHYAEGKVPKIKMALESILH 890	
QY	903 RRFTHQSDVWSYGVTVWELMTFGSKPYDGI PASDISSILEKGERLPQPPICTIDVYIMV 962	

DB	891 RIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASDISSILEKGERLPQPPICTIDVYIMV 950	
QY	963 KCMWIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSILLEDDMG 1021	
DB	951 KCMWIDADSRPKPRFRELILFESKWARDPQRYLVITQGDHMLPSPSTDSNFYRALMDEDEME 1010	
QY	1022 DLVDAEYLVPQOQFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSPL 1081	
DB	1011 DVDADEYLTPOQGF-----NSPST-----SRTPL 1036	
QY	1082 APSEAGSDVFDGDLGMGAAGLQSLPTHDPSPQLQRYSEDPTVPLPSET--DGYVAPLTC 1139	
DB	1037 LSSLSATS-----NSTVACINRNGSCRVEDAFQRYSSDPTGAVTEDNIDDAFL----- 1087	
QY	1140 SQPEVNVQDVPYRPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVEN 1199	
DB	1088 -VPVEYVNO-SVPKRPAGSVQNPVYHNPQHP-----APGRDLHYQN--PHSNAVGN 1135	
QY	1200 PLYL-TPQGGAAAPQHPPPAFPAFONLYWQD-----DP-----PERGAPPST 1242	
DB	1136 PEYLNTAQ-----PTCLSSGFNSPALWIKGSHQMSLDNPDYQDDFFPKETKPNGI 1186	
QY	1243 FKGTPTAENPEYGLDVP 1260	
DB	1187 FKQ-PTAENAEYLRVAPP 1203	
RESULT 4		
QYH40	PRELIMINARY; PRT; 1165 AA.	
ID	Q9YH40;	
AC	Q9YH40;	
DT	01-MAY-1999 (TrEMBLrel. 10, Created)	
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Receptor tyrosine kinase proto-oncogene.	
GN	XMRK.	
OS	Xiphophorus xiphidium.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;	
OC	Cyprinodontiformes; Poeciliidae; Xiphophorus.	
OX	NCBI_TaxID=8086;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RX	MEDLINE=98241172; PubMed=9582016;	
RA	Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,	
RA	Altshmidt J., Schartl M.;	
RT	"Activation of the Xmrk proto-oncogene of Xiphophorus by	
RT	overexpression and mutational alterations.";	
RL	Oncogene 16:1681-1690(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RIO PURIFICATION;	
RA	Schartl M.;	
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; U53471; AAD10500.2; ..	
DR	HSSP; P11362; 1FGK.	
DR	InterPro; IPR000345; CytC_heme_bind.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L_domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrcK; 1.	
DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.	
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.	

Best Local Similarity 46.9%; Pred. No. 1.6e-195; Matches 538; Conservative 172; Mismatches 351; Indels 85; Gaps 26;	
QY 161 LCYQDTILWKDIFKKNQALALTLIDNRSRACHPCSPMKSGRSGESSEDQSLTRTV 220	
DB 3 LCFADTHHQDVRPNMASNFTLVPTNGSSGGRCHKCTG-RCWGPTENHQTLTKTV 61	
QY 221 AGGC-ARCKGRLPTDCHEOAGCTGPKHSCLACLFPHNSGICELHCP-ALVFNFTV 278	
DB 62 AEQDRCYGVVSDCHREACGGCGPKDTPACMFNDSGACVTCQPTFVNPYTF 121	
QY 279 SFMLRVPKVASHLERYTFGASCTVACPNYLSTDVGSCTLVCPLNHOEVTAEDGTQCE 338	
DB 122 -----OLEHNNAKYTYGACVKKCHNFV-VDSSCVACRACPSKMEV-SENGIKMK 172	
QY 339 KSKPCARVCYGLGMEHLREVAVTSANIOEFAGCKIFGSLAFIPESFDGDPASNTAPL 398	
DB 173 PCTDLCPCACDGGTGSLSVSAOTVDSNIDKFNCTKINGNLIFLVGTHGDPYHTIAAI 232	
QY 399 OPEQLQVFETLEETGYLISAWPDSLPLSVFQNLQVIRILHNGAYSLTQGLGISW 458	
DB 233 NPEKLNIFQTVREITGYLNIQSWPENMTDFRVFSLNVTIGGRALYSGLSLLILKQGIT 292	
QY 459 IGLRSLRLGSLALIHNNTHLCFVHTVPWQDLFRNPHQALHTANRPEDECVGGLACH 518	
DB 293 LQFQSLKQISAGNIVITNSNLCCYHTVNTWTSLSFTPSQKTVIHRNKAENCTADGMVCN 352	
QY 519 OLCARGHGWGPTOCVNCQFLRGQECVEECRVLOGLPREYVNAHCLPCHPEQCP-ON 577	
DB 353 ELCSDDGCGPDDCLSKCRIRGTCIESNLVDGEFRFANGSVCMEDPQCEKMD 412	
QY 578 GSVTCFGEADOCVCAHYKDPFCVAPCPGVKPDLSYMPWKFPDEGACQCPINCT 637	
DB 413 NMITCYGCPDCTKCFHFQDGNVCEKCPDGLQANSF--IFKYADEDECHPCPNCT 470	
QY 638 HSCVDLDKGC-----PAORASPLTSIVSAVV-GILLVVLGVVFGILIKRRQ 685	
DB 471 QGCRGPASHDCIYYPTWRTSLPQHAR-TPL--IAAGVIGGLFIIVIMGLTEAVVRRKS 527	
QY 686 OKIRYWRRLLOETVELPLTPSGAMPNQAMRILKETLRKVKLGSAGTGVYKGM 745	
DB 528 IK-KRAURRLF-ETELVEPLTPSGTAPNQALRILKETLRKVVLGSGAGTGVYKGM 585	
QY 746 IPDGENVIPVAIKVLENTSPKANKEILDEAYVMAGVSPVSVRLGLICLTSTVOLVTO 805	
DB 586 VPEGETVPIPAIKILNETTGPKANVFMDALIMASMDHPLVRLGLVCLFTQLVTO 645	
QY 806 LMPYGLLDHVRNRLGSDQLLNCWQIAKMSYLEDVRLVLRDLAARNVLKSPNHV 865	
DB 646 LMPHGLLDYVHEHKDNIGSOLLNCWQIAKGMYLEERLVRDLAARNVLKSPNHV 705	
QY 866 KITDFGLARLLDIDETVHADGKVPKIMMALESILRRFTHOSDWSYGVTVWELMTFG 925	
DB 706 KITDFGLARLLDIDETVHADGKVPKIMMALESILRRFTHOSDWSYGVTVWELMTFG 765	
QY 926 AKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVKCMWIDSECRPFRELVSFSSRM 985	
DB 766 GKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVKCMWIDSECRPFRELVSFSSRM 825	
QY 986 ARDPORFVVIQNE-LGPASPLDSTFYRSLLEDMDGLVDAAEYLVPOQGFCDPAPG 1044	
DB 826 ARDPORYLVIOGDDRMKLSPNDSKFFQNLLEDDELDMDMAEYLVLP-QAFNIPPIYT 884	
QY 1045 AGGMVHRRSSRSTSGGDLTLGLEPSEEEAPRS--PLAP-SEGAGSDVFDGLGMGNA 1101	
DB 885 SSTRIDSNRNQVYRDGGVAEQGV-PMPYRAGCIIPAPVAQGATAEIFEDTCCNGTL 943	
QY 1102 KGLOSLPHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPOEYVNPQDVRPQ 1154	
DB 944 RKQVATLAKEDSSSTQRYSDPTVFIPERVIRCELDEDEGYMTPMRDKPTDYLNPVEENPF 1003	
QY 1155 PPSREGPLPAA-RPAGATLERAKTLSPGKNVVKDVF-----AFGGAIVENPEYLTPT 1205	
DB 1004 VSRKNGDLQAVDNPETHN-----APNGOPKADEYVNEPLYLNTFANTLENAEYL-- 1054	
QY 1206 OCGAAPQHPHPAFSPADNLVYWDODPERCA--PPSTFKGTPT-----A 1249	
DB 1055 -----KNLPEKAKAFDNDPDYMNHSUPPRSTLQHPDYQYSTKYFYKQNGRIRPIVA 1108	
QY 1250 ENPEYL 1255	
DB 1109 ENPEYL 1114	
RESULT 6	
ID P79754 PRELIMINARY; PRT: 1328 AA.	
AC P79754;	
DT 01-MAY-1997 (TREMBLrel. 03, Created)	
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)	
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE Erbb3.	
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;	
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;	
OC Tetraodontidae; Takifugu.	
OX NCBI_TaxID=31033;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=99177347; PubMed=10077531;	
RA Gellner K., Brenner S.;	
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu	
rubripes."	
RL Genome Res. 9:251-258(1999).	
DR EMBL; AF056116; AAC34391.1; --	
DR HSSP; P11362; IFKG.	
DR InterPro; IPR000494; EGFR_L domain.	
DR InterPro; IPR000719; Euk_pkinase.	
DR InterPro; IPR002174; Furin-like.	
DR InterPro; IPR001245; Tyr_pkinase.	
DR Pfam; PF00757; Furin-like; 1.	
DR Pfam; PF00069; pkinase; 1.	
DR Pfam; PF01030; Recep_L domain; 2.	
DR ProDom; PD000001; Euk_pkinase; 1.	
DR SMART; SM00261; FU; 3.	
DR SMART; SM00219; TyrKc; 1.	
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.	
KW ATP-binding; Transferase.	
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;	
Query Match 33.4%; Score 2284; DB 13; Length 1328;	
Best Local Similarity 40.3%; Pred. No. 1.3e-163;	
Matches 521; Conservative 157; Mismatches 417; Indels 198; Gaps 33;	
QY 9 WGLLALALPP--GAASQTQ---VCTGDMKRLPASPETHLDMLRLHYQGCQVVOGNLEL 62	
DB 4 WRLILMCVASRLRAASSTQAEVCPGTGNLSSGTQENQYNLNKDRYKGCIEIMGNLEI 63	
QY 63 TYLPTNASLSFLDQEQVYVLIHQNQVLPORLRIVRGTQLPFDNYALVLDNGDP 122	
DB 64 TQIESNWDPSFLKTIREVTGYVLIAMNHQEIPLQGLRVIRGNSLYERRFALSFLN--- 120	
QY 123 LNNTPVTGASPGGLRELQRLSLTEILKGVLIQNPOLCYQDTILWKDIFHKNQALAT 182	
DB 121 ----YPKDQ--PSGNLQGLMLNLTEILDGQVLIINNKYLRYGPWVYWRDII-RNNDAPIE 173	
QY 183 LIDTNRSRACHPCSPMKSGRSGESSEDQSLTRTVCAAGC-ARCKGRLPTDCCHCEQA 241	
DB 174 IQFNGERGVCV---KSC-GNYCNGFGKQCOQLTKTVCAPOCNDRCFGTSPRDCCHIECA 229	
QY 242 AGCTGPKHSDCLACLFHNSHGICELHCP-ALVFNFTVSWLRVPKVASHLERYTFGAS 300	
DB 230 AGCKGPLDTDFACRLFNDSGACVPQCPTLIYNKQTF-----QMETPNNAKYQYGS 282	

Qy	301	CVTACPNYLSTDVSGSTLCVPLHNQEVTAEDGTQR--CEKCSKPCARVCYGLGHEHURREV	359
Db	283	CVSQCPHFV--VDGSSCVSVCPDKMEV--ERGSQRQCELCSGLCPKVCBGTGAE--QR	336
Qy	360	RAVTSANTQEFAGCKKIFGSLAFIPESDGDGPASNTAPLOEQLOQVPETLEEITGLYLVIS	419
Db	337	QTVDSNIDSPINCTKIQSLHFLVTGLGDDFKNVPLDAKKLEVFRTVREITDILNIQ	396
Qy	420	AWPDSLPDLSPQNLQVIRGRIHLNGAYSLTLOGIGISWLGRLSRLRELGSGLAIHNNHTH	479
Db	397	SWPKELNDLSVFSLSLTIQGRSLFKRFSLMVMRIPTLTSLGRLSURLSIDGSVVISQNAH	456
Qy	480	LCFVHTVPWDOLFNPNH--QALLHTANRDEPCVCEGLACHOLCARGHCWGPQTCOVNCS	538
Db	457	LCYHHTVNWTLQFGRSVRANSLSNRPMAECVADRGVCDPLCSDSGCWGPGPOCLUSR	516
Qy	539	QFLRGQCEBECRVLQGLPREYVNAHH--CLPCHPECPQNGSVTCFGEADQCACAHYK	597
Db	517	NYSRHGTCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTCKASCTGFGADECWACTKFR	576
Qy	598	DPPFCVARGSPGVKPDLSYMPIWKFPOBEGACQCPINCTHSCVDLDDKGCPCAEQASPL	657
Db	577	DGPYCMSSCPAGVN--DGEKGLIFXFPNREGCEPCQHNCTQCSGPGGLNDC---LEAARL	632
Qy	658	TSIVSAVVUGILLVVVLGVVF-----GILIKRQQKIRKVTYMERLLLOETVELPPLPSG	710
Db	633	TISSGQITGALGVAPAGLICLVFLFGMLYHRGLAIRRKRAMRYLESGESFEPLGP-G	691
Qy	711	AMPNQAQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKAN	770
Db	692	EKGTKVHARILKPSDLAKIKPLGSGVGTGVSXGFWIPEGETVKIPVAIKTIQDSSGRQTP	751
Qy	771	KEILDEAYVMAGVCSPIVSRLLGLCLTSTVOLVTQLMPYGCLLDHHVRENRRGLSGDILLN	830
Db	752	TEITDHLLSMGSLDHPYIVRLLGTCPGTCQLVLTQLSHSGSLLEHIRHQHTSLDPORLLN	811
Qy	831	WCMQIAKGMYSLEDRVLRVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEXHADGKV	890
Db	812	WCVOIAKGMYYLEHRVHKNLAARNILLKNDYQVOISDYGVADLLPYDDKKYVYSETKT	871
Qy	891	PIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQP	950
Db	872	PIKMALESILFRYTHQSDVMSYGVTVWEMMSFGAEPIYASVQOEVPVSLVEKGERLSQP	931
Qy	951	PICTIDVYIMVKCWMIDSECRPFRELVSEFSSWARDPQRFVVIQNEDLGPASPLDSTF	1010
Db	932	AICTIDVYIMVKCWMIDENIRPTFKELASDFTRMARDPPRYLVIRNEG-----980	
Qy	1011	YRSILLEDDMGDLVDAEYILVPOQGFPCPDAPGAGGMVHHRHSSSTRSGGDLTLGLE	1070
Db	981	-----EDSGMGEFL-----RGSER---GLLEADLE	1003
Qy	1071	PSREEAPRSLAPSEAGAGSDVFDGLNG--AAKGLQSLPTHDPSPLO-----R	1117
Db	1004	EDBEE-----GLGDRFATPSLQPSWMTSPSQINSYVMVTQLR	1042
Qy	1118	YSEDPTVLPSETDGYVAPLTCSPQ--EYVNO-----PDVRPQPPSPREGP	1162
Db	1043	YD-----FAVQOGHHGYLPMSPSPVDITRQLWYQSRSLSSVRTLDPDRAFRSSREAE	1096
Qy	1163	L--PAARPAGATLERAKTLSPGKNGVVKVFAFGGAVENPEYLTPOCGAAPQHPPPAFS	1220
Db	1097	LCEDGAQCAGI FRV-----FGSERGN-----POCG-----	1122
Qy	1221	PAFDNLYWQDPPPERGAPSPSTFKGTPTANPE	1253
Db	1123	-----QORKLSTASSPSSFKTWAADBEDE	1146

RESULT 7

Q9BIH9

ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.

AC Q9BIH9;

DT	01-JUN-2001	(TReMBLrel. 17, Created)
DT	01-JUN-2001	(TReMBLrel. 17, Last sequence update)
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)
DE		Putative epidermal growth factor receptor (fragment).
GN		EGFR.
OS		Anopheles gambiae (African malaria mosquito).
OC		Eukaryota, Metazoa; Arthropoda; Hexapoda; Insecta;
OC		Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culic
OC		Anopheles.
OX		NCBI_TaxID=7165;
RN		{1}
RP		SEQUENCE FROM N.A.
RC		STRAIN=SUA;
RA		Lycett G.J.;
RT		"Cloning, expression and localisation of the Anopheles gambiae
RT		epidermal growth factor receptor.";
RL		Submitted (NOV-2000) to the EMBL/GenBank/DBDJ databases.

Query Match 29.4%; Score 2011.5; DB 5; Length 1433;
Best Local Similarity 32.5%; Pred. No. 6.1e-143;
Matches 472; Conservative 194; Mismatches 396; Indels 391; Gaps 37;

	Qy	26	CTGTDMLRLPASPETHDMLRHLYGCGVVOGNLELYLPTNASISFLQDIOEVQSYVL	85
Dd	1	CIGTGMSVPANREYHYKMLRDYNTCTVVDGNLEITMIONITDLNFLOHIREVTGYVL	60	

RESULT 7

Q9BIH9

ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.

AC Q9BIH9;

```
QY 430 VFQNLQVIRGILHNGAY-SLTQGLGISWLGRLSRLGSLALIHNNTHLCFVHTVPW 488
Db 384 YERNLEVVGRQLKENLFPASVIVKTSLSKLSKLRVNSGSIIVILENSDLCFVEDIDM 443
QY 489 DOLFNPQALHTANRDEVCVSGELACHQICARGHCWGPQTQVCNCSQFLRQOECVE 548
Db 444 SEIKSSDHEVMVQNRNATEHEGMECEQCSKAGCWGKGPEQCECKNVKYGKCLID 503
QY 549 ECRVLQGLPREY-VNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVAPC 607
Db 504 SKC---SLPRLYSVDSKTCGDCHQECQD-----FCYGNEDNCGSMNVKQGRFCVAECP 555
QY 608 -----BEG-----SGLVPLDLSYMPIMWPKPP-- 624
Db 556 TTTHAMNGTGINCHKTCVCGCRPRDTIAPDGCISCDAKIIIGSDAKIERCLMKDESCPDGY 615
QY 625 -----BEG-----SGLVPLDLSYMPIMWPKPP-- 627
Db 616 YSDYVLQEEGPLKQLSGKAVCRKCHPRCKKCTGYGPFHQECQCTGYKKGEQCECPQD 675
QY 628 -----ACQCPINCT-----HSCVDL-----DD-----KGCPEAQ--- 652
Db 676 FYANETRCLPCHQECRCCHGLGDHHECRNLKLFEGDPYDNATFTCVSNCPASHPYK 735
QY 653 -----RASPLTSIVSAVVGILLVVLGVVFGI---LKKRQOK 687
Db 736 RFPQBAKIGPYCSADSMQSLRIEPTQVKIVMSVMAILLVVGFAVFLSRHKNK 795
QY 688 IRKYTHRRLLQETELVEPTPGAMPNQAMRILKETELRKVKVLGSGAFGTVYGIWIP 747
Db 796 KDAVKMTALACEDSEPLRPSNVGNLTKLRIKEAEIRRGVGLMGAGFVRFGVWMP 855
QY 748 DGENVKIPVAIKVLRENTSPKANKEITLDEAYVMAGVSPVSRLLGICLTSTVQLVTQLM 807
Db 856 EGESVKIPVAIKVLEMSGESKEFLEAYIMASVEHPNLKLLKLAACVMTSQMMLITQLM 915
QY 808 PYGCLLDHVRNRRGSLQDLNWCMIQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKI 867
Db 916 PLGCLLDYVRNKKIGSKALLNWSQIARGMAYLEERLVRDLAARNVLVQTSCVKI 975
QY 868 TDFGLARLLDIDETEHADGGKVPKWMALSIILRRFTHQSDVWSYGVTVWELMTFGAK 927
Db 976 TVFGLAKLLDFFSDYRAAGGKMPIKWLALCIRHRVFTSKDVMVAFGITIWEILTYGAR 1035
QY 928 PYDGPAREIPDLLEKGERLPOPPICTIDVYIMVWKWIMIDSECRPFRELVSFESMAR 987
Db 1036 PHENVPAKDVPELIEIGHKLPOPDICSLDVCYLILSCWVLDADARPTFKQLAETFAEKAR 1095
QY 988 DPQRFVIONEDLGASPLDSTFYRSLLEDDEDDMDGLV----- 1024
Db 1096 DPGRYLMI-----PGDKFMRPLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNYD 1146
QY 1025 -----DABEYLVPQGFPCPDPAFCAGGMVHRRHSSSTRSGGDLTLGLFSEEEAPR 1078
Db 1147 VPSITAEITDYLQKTRPSIMLPGPSA-----VEFS-DEMPK 1182
QY 1079 S-----PLAP---SEGAGSDVFDGLGMAAKLQSLPHTDPSPLQRYSEDPTVPLPSE 1129
Db 1183 SLRYCKDPLKPDDETDGHEV-----GVGGIR-----LNLPID 1216
QY 1130 TDGYVAPLTCSPQEVYVNDVRPOPSPREGFLPAARPAATLERAKTLSPGKNGVWKD 1189
Db 1217 EDDYLMPTCQSQ---NQS-----TFG---YMD 1237
QY 1190 VFAFGAVENPEYL-----TPQGGNAPQHPHPPAFSPAFDNLVYWDQDPPERGAPP 1240
Db 1238 LIGVPASVDNPEYLMGSTQIAGLAQSGMG--PHTPP-----PP 1274
QY 1241 STFKGTPTAENPE 1253
Db 1275 NTPNGMPTHQHSQ 1287
```

RESULT 8

```
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347B2D030C CRC64;
```

Query Match 25.9%; Score 1770.5; DB 4; Length 419;
Best Local Similarity 85.0%; Pred. No. 1.7e-125;
Matches 340; Conservative 7; Mismatches 40; Indels 13; Gaps 3;

```
QY 1 MELAALCRWGLLLALLPGGAASQVCTGTDMLKRLPASPEHLMLRHLYQGCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPGGAASQVCTGTDMLKRLPASPEHLMLRHLYQGCQVVOGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIQAHNQVQPLQRLIRVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGYVLIQAHNQVQPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLTQRPOLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLTQRPOLCYQDTILWKDIFHKQNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVACGACARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSWFLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSWFLRVPKVSASHLERYTFGAS 294
QY 301 CVTACPNYLSLTDVSCCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGHEHLREVR 360
Db 295 CVTACPNYLSLTDVSCCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLGHEHLREVR 349
QY 361 AVTSANIOEFAGCKKIFGSLAFPLSPESFDGDPASNTAPLOP 400
Db 350 AAVPVLNRMQPG--PAHPVLSPFLRSPDLVSFAFSLPLAP 387
```

RESULT 9

```
Q8R2X1
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
```

OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Apr-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC027080; AAH27080.1; -;
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 3.3e-123; Indels 0; Gaps 0;
Matches 323; Conservative 11; Mismatches 33;

QY 895 MALESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 954
DB 1 MALESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPIC 60

QY 955 IDVYIMVWCMWIDSECRPRFRELVSFSESRMARDPQRFVIONEDLGPASPLDSTFYRSL 1014
DB 61 IDVYIMVWCMWIDSECRPRFRELVSFSESRMARDPQRFVIONEDLGPSSPMDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEEYLVPOGGFFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEEYLVPOGGFFSPDPALGTGTARRHRRSSSARSGGELTLGLEPSEE 180

QY 1075 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSPETDGV 1134
DB 181 EPRSLAPSEGAGSDVFDGDLVAVGTVKGLQSLSPHLSPLQRYSEDPTLPPETDGV 240

QY 1135 APLTCSPOEYVQNPQVPPPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAG 1194
DB 241 APLACSPQEYVQNPQVPPPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAG 300

QY 1195 GAVENPEYLTPOGGAPOPHPPAPSPAFDNLVYWDQPPERAGPSTFKGTPTAENPEY 1254
DB 301 GAVENPEYLAAGTASQPHSPAPSPAFDNLVYWDQNSSEGGPPSTFEGPTAENPEY 360

QY 1255 LGLDVPV 1261
DB 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.;
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -;
DR HSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.

DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TVR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFEID63 CRC64;

Query Match 25.1%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 2.5e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 575 PONGSVTCFPEADQCVACAHYKDPFPCVAPCPSPGVKPDLSYMPIWKFDPDEGACQPCPI 634
DB 141 PEETATPKTP--DHCMKCAHFIDGPHCVKACPAVGLGENDTL-VNKYADANAVCOLCHP 197

QY 635 NCTHSCVDLDDKGAEPQASPLTSTISAVV--GILLVVVLGVVFGVLLIKRQOKIRKVTM 693
DB 198 NCTRCCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVRKRTL 253

QY 694 RRLQETELVELPTSPGAMPNOAMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVK 753
DB 254 RELLOERELVELPTSPGEAPNOAHLRLKETEFKVKVLGSGAFGVYKGLWIPESKVK 313

QY 754 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPYSRLGICLTSTVQLVTQLMPYGCLL 813
DB 314 IPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCLL 373

QY 814 DHVRENRLGSGODLLNMCWIAKMSYLEDVLRVLRDLAARNLVKSPNHVKITDFGLA 873
DB 374 DYIREHKONIGSYQLNMCVQIAKGMNLYLEERLVRDLAARNLVKTPQHVKITDFGLA 433

QY 874 RLDDIDETEHADGGKVPKWMALSIILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGP 933
DB 434 KLGADAEYHAEKGVPIKWMALSIILHRIYTHQSDVMSYGVTVVWELMTFGSKPYDGP 493

QY 934 AREIPDLLEKGERLPQPPICITIDVTMVMKWMIDSECRPRFRELVSFSESRMARDPQRFV 993
DB 494 ASETSVLEKGERLPQPPICITIDVTMVMKWMIDADSPPKRELIAEFSKWARDPPRYL 553

QY 994 VIQ--NEDLGPASPLDSTFYRSLLEDMDGLVDAEEYLVPOGGFFCDPAPGAGMVHHR 1052
DB 554 VIQGERVHRLSPSTDSKFRITLMEBEDMEDIVDADEYLVPHOGFF-----598

QY 1053 HRSSSTRSGGDLTLGLEPSEEAAPRSL-----APSEGAGSDVFDGDLGMAAGLQSL 1107
DB 599 -NSPST-----SRTPLLSLSATSNSATNCID-----RNGQGH 631

QY 1108 PTHDPSPLQRYSEDPTVLPSPET--DGYVAPLTCSPQPEYVQNPQVPPSPREGPLPA 1165
DB 632 PVREDSFVQRYSSDPTGNFLBESIDGFL-----PAPEYVQ--LMPKKPS-----675

QY 1166 ARPAGATLERAKTILSPGKNGVVKDVF-----AFGGAIVENPEYL 1203
DB 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE V-ERBB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,

RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -;
DR HSSP; P11362; IFGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 6390 MM; C6D9CBA7ADF725E1 CRC64;

Query Match 25.1%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 2.5e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 584 GPEADQCACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 643
DB 1 GP-DHCKCAHFIDGPHCVKACPACVAGENDTL-VKYNADANAVCOLCHPNCTRCCKGP 57

QY 644 DKGCPAQRASPLTISVNAV-GILLVVLGVVFGILIKRQOKIRKYMRLLOTEL 702
DB 58 GLEGCP--NGSKTPSIAGVGGGLVWGLGIGLYLRR-HIVRKRTLRLLOREL 113

QY 703 VEPLTPSGAMPQAOIRLKETELRKVKVLGSGAGFTVYKGIWIPGENVKIPVALKVR 762
DB 114 VEPLTPSGAPQAHRLIKETEFKKVKVLGSGAGFTVYKGLWIPGEKVKIPVALKEUR 173

QY 763 ENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVOLVQLMPYGCLLDHRVNRGR 822
DB 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVOLITQLMPYGCLLDVIRESKON 233

QY 823 LGSODLLNWCMOIAKMSVLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSYLLNWCVOIAKMNLYEBRRLLVHRDLAARNVLVKTPOHVKITDFGLAKLLGADEKE 293

QY 883 YHADGKVKPKWMALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLE 942
DB 294 YHAEGKVKPKWMALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLE 353

QY 943 KGERLPQPICTIDVYIMVWKCMIDSECRPRFRELVSFERSMARDPQRFVVIQ-NEDLG 1001
DB 354 KGERLPQPPICTIDVYIMVWKCMIDADSRPKFRELIAEFSKWARDPPRYLVVIQDERMH 413

QY 1002 PASPLDSTYRSILLEDDMGDLVDAEYLVPOQGFCDPAPGAGGVVHRRSSSTRSG 1061
DB 414 LPSPTDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

QY 1062 GGDLTGLGPESEEEAPRSL-----APSECAQSDVFDGLGMAKGLQSLPHTDPSPLQ 1116
DB 455 -----SRTLLSLSATSNNSATNCID-----RNGQGHVPREDSFVQ 491

QY 1117 RYSEPTVPLPSET--DGVVAPLTCSPQYEVNQPDVRPQPPSPREGPLPAARAGATLE 1174
DB 492 RYSSDPTGNFLEESIDDGFL-----PAPEYVYQ--LMPKPS----- 526

QY 1175 RAKTLSPGKGVVVKDVF-----AFGGAIVENPEYL 1203
DB 527 ----TAMVQNIYNNISLTAIKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;

DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PP3859.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -;
DR InterPro: IPR002048; Euk_pkinase.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00108; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MM; 034397FF3F272D28C CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 5.5e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 954
DB 1 MALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICT 60

QY 955 IDVYIMVWKCMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGASPLDSTFYRSL 1014
DB 61 IDVYIMVWKCMIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGASPLDSTFYRSL 120

QY 1015 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEE 1074
DB 121 LEDDDMGDLVDAEYLVPOQGFCDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSEE 180

QY 1075 EAPRSLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1134
DB 181 EAPRSLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

QY 1135 APLTCSPOEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVVVKDVPFAG 1194
DB 241 APLTCSPOEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTLSPGKGVVVKDVPFAG 300

QY 1195 GAVENPEYLTPOGGAAPQ-----HPPPA---FSPAFDNL 1226
DB 301 GAVENPEYLTPOGGAALSPHTLLPSAQPSSTTITGTRTHQSGGLHPAPSKGHLRQRTQST 360

QY 1227 YWVD-QDPPER-----GAPSTFKGTPTAEN 1251
DB 361 WYWTQCEPEQGVRRSPDVSSGREGLTSAIKRWEGPPTTSRGTCARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN Gag, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.

OC Vituses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskun A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -;
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zF-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR PRODOM; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 109320 MW; 3C5AED791EAE95CE CRC64;

Query Match 24.2%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 4.1e-116;
Matches 358; Conservative 73; Mismatches 149; Indels 119; Gaps 18;

QY 547 VECCRVLOGLPRE-VYNNAR-HCLP-----CHPQCQ 574
DB 354 IEKQESYLLAFAYINIRKGINIFHWSKLLMKVADLRMIGAYHASRFLHMKVECPTELS 413
QY 575 PONGSVTCFGEADQCVACAHYKDPFCVACRCPKDLSPYMPWKFPDEEGACQCPPI 634
DB 414 PQE-----VGP--DHCNKAHFIDGPHCVKACPAVLGENDTL-VWKTADANAVCQLCHP 465
QY 635 NCTHSCVDLDDKGPACQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTM 693
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGLGIGLYLRRR-HIVRKRTL 521
QY 694 RLLQETLEVEPLTPSGAMPNQAMRILKETELRKVKVLSGAGTGYVKGWIIPDGENVK 753
DB 522 RLLQERELVEPLTPSGEAPNQAHRLIKETEFKKVKVLSGAGTGYVKGWIIPDGEKVT 581
QY 754 IPVAIKVLRNTPSKANKEILDEAYVMAGVSGPVVSRLLGICLTSTVOLVTQMPYGCGLL 813
DB 582 IPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTQMPYGCGLL 641
QY 814 DHVRENRLGSDLLNWCQIAKMSYLEDRVLVHRDLAARNVLKSPNHVKITDFGLA 873
DB 642 DYIREHKNIGSQYLLNWCQIAKMSYLEDRVLVHRDLAARNVLKSPNHVKITDFGLA 701
QY 874 RLLDIDETEYHAGDKGKPIKMALESILRRPFTQSDVWSVGVTVWELMTFGAKPYDGIP 933
DB 702 KQLGADKEHYHAEGGKVPKMALESILHRYTHQSDVWSVGVTVWELMTFGSKPYDGIP 761
QY 934 AREIPDLLEKGERLPQPPICTIDVYIMIMVKCWMIDSECRPFRELVSFESRMARDPORFV 993
DB 762 ASEISSVLEKGERLPQPPICTIDVYIMIMVKCWMIDSECRPFRELVSFESRMARDPORFV 993
DB 762 ASEISSVLEKGERLPQPPICTIDVYIMIMVKCWMIDSECRPFRELVSFESRMARDPORFV 993

QY 994 VIQ-NEDLGASPLDSTFYRSLLDEDDMGDLVDABEYLVPQOQFFCFDPPAGGVMVHHR 1052
DB 822 VIQGERMHLPSPTDSKFTYRLMEBEDMEDIVDADLYLPHQGGF-----866
QY 1053 HRSSSTRSGGDLTLGLPESEEEAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPHTDP 1112
DB 867 -NSPST-----SRTPLLSLSATS-----NSATKCIDRNGGH-- 898
QY 1113 SPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYVNPQDVRPOPSPREGPLPAARAGAT 1172
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQONQIYINISLT 936
QY 1173 -LERAKTILSPGKGVVVDVFAFGGAVENPEYL 1203
DB 937 AISKLPMDSRVQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468
ID Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;

Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 7.8e-116;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GPEADQCVACAHYKDPFCVACRCPKDLSPYMPWKFPDEEGACQCPINCTHSCVDL 643
DB 1 GP--DHCNKAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCRTGCKGP 57
QY 644 DDKCQPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTHRLLQETEL 702
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGLGIGLYLRRR-HIVRKRTLRLQLEREL 113
QY 703 VEPLTPSGAMPNQAMRILKETELRKVKVLSGAGTGYVKGWIIPDGENVKIPVAIKVL 762
DB 114 VEPLTPSGEAPNQAHRLIKETEFKKVKVLSGAGTGYVKGWIIPDGEKVTIPVAIKEL 173
QY 763 ENTSPKANKEILDEAYVMAGVSGPVVSRLLGICLTSTVOLVTQMPYGCGLLDHVRNRR 822
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLVTQMPYGCGLLDYIREHKN 233
QY 823 LGSODLLNWCQIAKMSYLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
DB 234 IGSQYLLNWCQIAKMSYLEDRVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882

QY 883 YHAGGKVPKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 942
 DB 294 YHAEKGKVPKMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIAREISSVLE 353
 QY 943 KGERLPQPICTIDVYIMVWCWMDISCRFRFRELVSFESFMRDQRFVVIQ-NEDLG 1001
 DB 354 KGERLPQPICTIDVYIMVWCWMSDADSRPKFRELIAEFSSMRDPPRYLVIOGDERMH 413
 QY 1002 PASPLDSTFYRSLDEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSG 1061
 DB 414 LPSPTDSFYRSLDEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSG 1061
 QY 1062 GGDLTGLGLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGLQSLTPHDPSPLQ 1116
 DB 455 -----SRTPLLSSLSATSNNNSATNCIDRNGG-----H----- 481
 QY 1117 RYSEDPVPLPSETDGYVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-IER 1175
 DB 482 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQIOIYISILTAISK 523
 QY 1176 AKTLPSPGKNGVVKDVFAFGGAVENPEYL 1203
 DB 524 LPMDSRYQN-----SHSTAVDNPEYL 544

RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.
 AC Q9WVF5;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
 DE isoform 3).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maible N.J.;
 RA "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVITAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J., Sinclair C.S., Pearball R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maible N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staab F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzei J., Momberts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S., Havaehizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 405:685-690(2001).
 DR EMBL; AF124513; AAD44149.1; --
 DR EMBL; AF275366; AAG28047.1; --
 DR EMBL; AF275364; AAG28047.1; JOINED.
 DR EMBL; AF275365; AAG28047.1; JOINED.
 DR EMBL; AK004944; BAB23688.1; --
 DR EMBL; AK004883; BAB23641.1; --
 DR EMBL; AK004911; BAB23662.1; --
 DR MGI; MGI:95294; Egfr.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR002174; Furin-Like.
 DR Pfam; PF00757; Furin-Like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR SMART; SM00261; FU; 3.
 KW Receptor.
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.7%; Score 1481.5; DB 11; Length 655;
 Best Local Similarity 43.3%; Pred. No. 2.4e-103;
 Matches 279; Conservative 100; Mismatches 237; Indels 29; Gaps 8;
 QY 11 LLLALLPPGAA--STOVCTGTDMKRLRLPASPEHLDMLRLHLYOGVQVQGNLELTVLPYN 68
 DB 14 LTLALCAAGGAALEKKVCGQTSNRLTQGTGFEDHFLSLQRMYNVNCVGLNLEITTVQRN 73
 QY 69 ASLSFLQDIQEVQGVYLIHNRQVPLQRLIRVRGTQLFEDNYLAVALDNGDPLNNTTP 128
 DB 74 YLSFLKTIQEVAGVYLIHNRQVPLQRLIRVRGTQLFEDNYLAVALDNGDPLNNTTP 124
 QY 129 VTGASPGRLRELQRLSLTEILKGGVLIQORNOLCYQDITLWKDI-----PHKNQLALTLI 184
 DB 125 -YGNRTGLRELPMRLNLEILIGAVRFNSNPLCNMDTIQMRDIVQNVFMNSMDL--- 180
 QY 185 DTRNSRACHPCSPCKSGRCNGESSEDCQSITRTVCAGGCA-RCKGPLPTDCHEOCAAG 243
 DB 181 -QSHPPSCPKCDPCPCNGSCWGGGEECQKTKIICAQCSHRCGRSPDCCHNQCAAG 239
 QY 244 CTGPKHSDCLACLHFNHSGICELHCPALVFNFTVSFWLRVPKVSASHLERYTFGASCVT 303
 DB 240 CTGPRESDECLVCQKQFDEATCKDTCPPLMLYNPTTY-----QMDVNPBGKYSFGATCVK 293
 QY 304 ACPVNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLRVRVAVT 363
 DB 294 KCPRYVVTDHGSCVRACGPDYEV-EEDGIRKCKKCDGPKCKVNGIGIGFKDTLSIN 352
 QY 364 SANTQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQEQLOVFTLEETITGLYTSAWPD 423
 DB 353 ATNIGHFKYCTAISGDHLILPVAFKDSFTPTPLDPRELBILKTVKEITGELLQAWPD 412
 QY 424 SLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRSLRSLRSLRSLRSLRSLR 483
 DB 413 NWTDLHAFENLEIRGRTRKQHQFSLAVVGLNITSLGLRSLRSLRSLRSLRSLRSLRSLR 472
 QY 484 HTVPWDQLFRPHQALLHTANRPEDECVEGLACHQLCARHCWCWGPGPTQCVCNQFLRG 543
 DB 473 NTINWKKLFGTPNQKTKIMNRAEKDKAVNHVCNPLCSSEGCWGPEDRCDVSCQNVSRG 532
 QY 544 QCVVECEVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCV 603
 DB 533 RECVEKCNILGEPRFVENSECTQCHPECLPQANNITCTGRGPNCTQCAHYIDGPHCV 592
 QY 604 ARCPGKVPKDLSSYMPKIPWFPDEEGACQPCINCTHSCVDLDDKGC 648
 DB 593 KTCPCAGIMGNNTL-VMKYDANNVCHLCHANCTYGCAGPGIQQC 636

Search completed: July 22, 2003, 09:00:55
Job time : 53.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-250-264-14
Perfection score: 6839
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : A_Geneseq_101002.*
- 1: /SID22/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.*
 - 2: /SID22/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.*
 - 3: /SID22/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.*
 - 4: /SID22/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.*
 - 5: /SID22/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.*
 - 6: /SID22/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.*
 - 7: /SID22/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.*
 - 8: /SID22/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.*
 - 9: /SID22/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.*
 - 10: /SID22/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.*
 - 11: /SID22/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.*
 - 12: /SID22/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.*
 - 13: /SID22/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.*
 - 14: /SID22/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.*
 - 15: /SID22/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.*
 - 16: /SID22/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.*
 - 17: /SID22/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.*
 - 18: /SID22/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.*
 - 19: /SID22/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.*
 - 20: /SID22/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.*
 - 21: /SID22/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.*
 - 22: /SID22/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.*
 - 23: /SID22/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6708	98.1	1255	21	Human heregulin 2
2	6708	98.1	1255	22	Human tyrosine kin
3	6708	98.1	1255	22	HER2 transgene pla
4	6708	98.1	1255	23	Human HER2 (ErBB2)
5	6702	98.0	1255	17	HER-2/neu protein.
6	6702	98.0	1255	20	Human HER-2/neu on
7	6702	98.0	1255	21	Human HER-2/neu pr
8	6702	98.0	1255	21	Amino acid sequenc
9	6702	98.0	1255	22	Human HER-2/neu pr
10	6702	98.0	1255	22	HER2/neu amino aci

11	6702	98.0	1255	23	AAE24067	Human Her-2 protei
12	6702	98.0	1255	23	AAE20479	Human Her-2/neu pr
13	6702	98.0	1255	23	AAE20479	Human Her-2/neu pr
14	6702	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6659	97.4	1433	14	AAE39568	Sequence of c-erbB
16	6538	95.6	1223	23	AAU98923	Human breast cance
17	6385	93.4	1200	21	AAE21208	Human HER-2/neu pr
18	5905.5	86.4	1256	21	AAE21199	Rat Her-2/neu pr
19	5905.5	86.4	1256	23	AAE21199	Rat Her-2/neu pr
20	5881.5	86.0	1256	21	AAE21206	Mouse Her-2/neu pr
21	5881.5	86.0	1256	22	AAE21206	Mouse Her-2/neu pr
22	5881.5	86.0	1256	23	AAE21206	Amino acid sequenc
23	4788	70.0	919	21	AAE21203	Mouse Her-2/neu on
24	4788	70.0	919	23	AAE21203	Human HER-2/neu fu
25	4031.5	58.9	920	23	AAE21148	Her-2/neu extracel
26	4031.5	58.9	926	23	AAE21152	Mouse Her-2/neu ex
27	3672	53.7	712	21	AAE21204	Human HER-2/neu fu
28	3672	53.7	712	21	AAE21204	Human HER-2/neu fu
29	3526	51.6	782	18	AAE21149	Her-2/neu extracel
30	3524	51.5	653	21	AAE21200	Her2-GM-CSF immuno
31	3524	51.5	653	23	AAE21200	Extracellular HER-
32	3486	51.0	645	22	AAE21145	Human Her-2/neu on
33	3486	51.0	645	22	AAE21145	Human ErbB2 oncopr
34	3421	50.0	645	22	AAE21153	Human ErbB2 extrac
35	3318	48.5	624	11	AAE44993	DC9scFv-erbB2EC fu
36	3116	45.6	1210	21	AAE19259	Amino acid sequenc
37	3116	45.6	1210	21	AAE19259	Extracellular port
38	3116	45.6	1210	23	AAE23019	Human EGF receptor
39	3116	45.6	1210	23	AAE23019	Human Her-1 protei
40	3114	45.5	1210	22	AAE50768	Human epidermal gr
41	3084	45.1	583	23	AAE68420	Amino acid sequenc
42	3084	45.1	587	23	AAE20483	Human protein for
43	3083	45.1	589	23	AAE20481	Human protein for
44	3083	45.1	600	23	AAE20484	Human protein for
45	3075	45.0	1210	23	AAE20482	Human epidermal gr

ALIGNMENTS

RESULT 1
AAE20482
ID AAY92620 standard; Protein; 1255 AA.
XX
AC AAY92620;
XX
DT 10-AUG-2000 (first entry)
XX
DE Human heregulin 2 (Her2).
XX
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
OS Homo sapiens.
XX
FH Key
FH Domain
FT Location/Qualifiers
FT 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT 174..323
FT Domain

FT Region /label= Cysteine_rich_domain
 210..224
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 250..264
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 324..483
 /label= Ligand_binding_domain
 325..339
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 369..383
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 465..479
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 484..623
 /label= Cysteine_rich_domain
 579..593
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 624..654
 /label= Transmembrane_domain
 632..652
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 653..667
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 655..1010
 /label= Tyrosine_kinase_domain
 661..675
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 695..709
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 710..730
 /label= insertion_region
 /note= "suitable for foreign epitope insertion"
 1011..1235
 /label= C-terminal_domain
 WO200020027-A2.
 13-APR-2000.
 05-OCT-1999; 99WO-DK00525.
 05-OCT-1998; 98DK-0001261.
 20-OCT-1998; 98US-0105011.
 (MEBI-) M & E BIOTECH AS.
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 Gautam A, Birk P, Karlsson G;
 WPI; 2000-349917/30.
 N-PSDB; AAA09455.
 Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer
 Claim 62; Page 193-198; 220pp; English.
 This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX
 SQ Sequence 1255 AA;
 Query Match 98.1%; Score 6708; DB 21; Length 1255;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;
 QY 1 MELAALCRWGLLLALLPFGAASQTCTGTDMLRLPASPTHLDMLRHLVYQGVVQGNL 60
 Db
 1 MELAALCRWGLLLALLPFGAASQTCTGTDMLRLPASPTHLDMLRHLVYQGVVQGNL 60
 QY 61 ELTYLPTNASLFLQDIOEVQGVYLIAHNOVRQVPLRLRIVRGTLQEDNALAVLDNG 120
 Db
 61 ELTYLPTNASLFLQDIOEVQGVYLIAHNOVRQVPLRLRIVRGTLQEDNALAVLDNG 120
 QY 121 DPLNNTTPTVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 Db
 121 DPLNNTTPTVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSKGRCMGESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240
 Db
 181 LTLIDTNRSRACHPCSKGRCMGESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSFWLRVPKVSASHLERTFFGAS 300
 Db
 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVFNFTVSFWLRVPKVSASHLERTFFGAS 300
 QY 301 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVR 360
 Db
 295 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKCSKPCARVCYGLGMEHLREVR 354
 QY 361 AVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETLEEITGYLYISA 420
 Db
 355 AVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLQPEQLQVFETLEEITGYLYISA 414
 QY 421 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNNTHL 480
 Db
 415 WPDLSLPDLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGSLALIHNNTHL 474
 QY 481 CFVHTVPMDQLFRNPQHALLHTANRPEDECVEGGLACHQLCARGHCWGPGPTQCVNCSQF 540
 Db
 475 CFVHTVPMDQLFRNPQHALLHTANRPEDECVEGGLACHQLCARGHCWGPGPTQCVNCSQF 534
 QY 541 LRQECVEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
 Db
 535 LRQECVEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
 QY 601 FCVARCPGVKPDLSYMPDWKFPDEEGACQPCPNCTHSCVDLDDKGCPEAQRASPLTSI 660
 Db
 595 FCVARCPGVKPDLSYMPDWKFPDEEGACQPCPNCTHSCVDLDDKGCPEAQRASPLTSI 654
 QY 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVBEPLTPSGAMPNOAMRI 720
 Db
 655 VSAVVGILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVBEPLTPSGAMPNOAMRI 714
 QY 721 LKETELRKVKVLSGAGFTGVYKGIWIPDGENVKVIPVAIKVLRENTSPKANKEILDEAYVM 780
 Db
 715 LKETELRKVKVLSGAGFTGVYKGIWIPDGENVKVIPVAIKVLRENTSPKANKEILDEAYVM 774

QY	781	AGVGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDLDLNNWCQIAKGS	840
Db	775	AGVGSPPYVRLIGICLTSTVQLVTQMPYGCCLLDHVRENRLGSGDLDLNNWCQIAKGS	834
QY	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWALLES	900
Db	835	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWALLES	894
QY	901	LRRRFTHQSDVMSYGVTTWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI	960
Db	895	LRRRFTHQSDVMSYGVTTWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMI	954
QY	961	MVKCMIDSECRPRELVSERPMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM	1020
Db	955	MVKCMIDSECRPRELVSERPMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM	1014
QY	1021	GDLDVAEEVLVPOGGFFCPDPAPGAGMWVHRRSSSTRSGGDLTLGLEPSEEAAPRSP	1080
Db	1015	GDLDVAEEVLVPOGGFFCPDPAPGAGMWVHRRSSSTRSGGDLTLGLEPSEEAAPRSP	1074
QY	1081	LAPSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCS	1140
Db	1075	LAPSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCS	1134
QY	1141	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPQKNGVVDVFAFGAVENP	1200
Db	1135	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPQKNGVVDVFAFGAVENP	1194
QY	1201	EYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	1260
Db	1195	EYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	1254
QY	1261	V 1261	
Db	1255	V 1255	
RESULT 2			
AAE12130			
ID	AAE12130	standard; Protein; 1255 AA.	
AC	AAE12130;		
XX			
DT	18-DEC-2001	(first entry)	
DE		Human tyrosine kinase-type receptor, HER-2.	
KW		Therapeutic compound; major histocompatibility complex; vaccine; antigenic peptide; MHC; immunoregulatory; immune response; HER-2;	
KW		adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;	
KW		antigen presenting cell; human; tyrosine kinase-type receptor.	
OS		Homo sapiens.	
XX			
EH	Key	Location/Qualifiers	
FT	Region	774..782	
FT		/note= "Antigenic epitope"	
XX			
XX	WO200168677-A2.		
XX	20-SEP-2001.		
XX			
XX	16-MAR-2001; 2001WO-US40328.		
XX			
XX	16-MAR-2000; 2000US-0527487.		
XX			
PA	(GENZ) GENZYME CORP.		
PI	Nicolette CA;		
XX			
XX	WPI; 2001-616284/71.		
DR	N-PSDB; AAD19731.		
XX			

PT		Novel synthetic therapeutic compound for inducing immune response and for use in adoptive immunotherapy, has enhanced binding to major histocompatibility molecules and enhanced immunoregulatory properties	
PT			
PT			
XX			
PS	Claim 4; Page 63-67; 69pp; English.		
XX		The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterised by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm translocation of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).	
SQ	Sequence 1255 AA;		
	Query Match	98.1%; Score 6708; DB 22; Length 1255;	
	Best Local Similarity	98.5%; Pred. No. 0;	
	Matches 1242; Conservative	2; Mismatches 11; Indels 6; Gaps 1;	
QY	1	MELAAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPETHLDMRLHLYQCGVQGNL	60
Db	1	MELAAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPETHLDMRLHLYQCGVQGNL	60
QY	61	ELTYLPTNASISFLQDIOEVQGYVLI AHNVQVFLQRLIRVGTQFSDNTALAVLDNG	120
Db	61	ELTYLPTNASISFLQDIOEVQGYVLI AHNVQVFLQRLIRVGTQFSDNTALAVLDNG	120
QY	121	DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQNPOLCVQDTILWKDIFHKNOLA	180
Db	121	DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQNPOLCVQDTILWKDIFHKNOLA	180
QY	181	LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC	240
Db	181	LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARCKGLPTDCCHEQC	240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS	294
Db	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS	294
QY	301	CVTACPNYLSLTDVGSCTLVCPPLHNOEVTAEQGTORCEKCKPCARVCYGLGMEHLREVR	360
Db	295	CVTACPNYLSLTDVGSCTLVCPPLHNOEVTAEQGTORCEKCKPCARVCYGLGMEHLREVR	354
QY	361	AVTSANIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFTLEETGYLYISA	420
Db	355	AVTSANIOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVFTLEETGYLYISA	414
QY	421	WPSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGLALIHNNTHL	480
Db	415	WPSLPLDLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGLALIHNNTHL	474
QY	481	CFVHTVPMWDLFRNPHQALLHTANRPEDECYEGGLACHOLCARHCWGPGTQCVCNCSOF	540
Db	475	CFVHTVPMWDLFRNPHQALLHTANRPEDECYEGGLACHOLCARHCWGPGTQCVCNCSOF	534
QY	541	LRGQECVVEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFPGPEAQCVACAHYKOPP	600
Db	535	LRGQECVVEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFPGPEAQCVACAHYKOPP	594
QY	601	FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCABORASPLTSI	660
Db	595	FCVARCPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCABORASPLTSI	654


```
QY 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 720
Db 655 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 780
Db 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 774
QY 781 AGVGSPPVSRLLGICLTSTVQLVTOALMPYGCILLDHVRENRRGLGSODLNLWCWQAKGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLVTOALMPYGCILLDHVRENRRGLGSODLNLWCWQAKGMS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 894
QY 901 LRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 960
Db 895 LRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 954
QY 961 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVFQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEEYLVFQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTDGYVAPLTCS 1134
QY 1141 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;
```

```
XX WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6708; DB 22; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLRASPETHLDMRLHLYQGCQVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLRASPETHLDMRLHLYQGCQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKPLPTDCCHEOC 240
Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACARCKPLPTDCCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICBLHCPALVFNFTVSFWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICBLHCPALVFNFTVSFWLRVPKVSASHLERYTFGAS 294
QY 301 CVTACPYNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVR 360
Db 295 CVTACPYNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVPFETLEEITGYLYISA 420
Db 355 AVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVPFETLEEITGYLYISA 414
QY 421 WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIHNNTHL 480
Db 415 WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLGSLALIHNNTHL 474
QY 481 CFVHTVPMWDLPRNPHQALLHTANPEDECEVGEGLACHQLCARGHCWGPGTQCVCNSOF 540
Db 475 CFVHTVPMWDLPRNPHQALLHTANPEDECEVGEGLACHQLCARGHCWGPGTQCVCNSOF 534
QY 541 LRQGECEBECEVLRQGLPREYVNNARHCLPCHPCQPNQSGSVTCFGEADOCVACAHYKDP 600
Db 535 LRQGECEBECEVLRQGLPREYVNNARHCLPCHPCQPNQSGSVTCFGEADOCVACAHYKDP 594
QY 601 FCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTSI 654
QY 661 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 720
Db 655 VSAVVGILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOQMRI 714
QY 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 780
Db 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKSILDEAYYM 774
QY 781 AGVGSPPVSRLLGICLTSTVQLVTOALMPYGCILLDHVRENRRGLGSODLNLWCWQAKGMS 840
Db 775 AGVGSPPVSRLLGICLTSTVQLVTOALMPYGCILLDHVRENRRGLGSODLNLWCWQAKGMS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESI 894
QY 901 LRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 960
Db 895 LRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPCTIDVYMI 954
QY 961 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVFQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEEYLVFQOGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDSPLQRYSEDPVPLPSTDGYVAPLTCS 1134
QY 1141 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGGAAQPPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;
```

Db 715 LKETELRKVKVLGSGAGFTYKGIWIPDGENVKI PVAIKVIRENTSPKANKEILDEAYVM 774
QY 781 AGVGSYPVSRLLIGICLTSTVQLVTQMPYGCCLLDHVRENRRGRSGODLLNWCNQIAKGS 840
Db 775 AGVGSYPVSRLLIGICLTSTVQLVTQMPYGCCLLDHVRENRRGRSGODLLNWCNQIAKGS 834
QY 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMKMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMKMALESI 894
QY 901 LRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGYVTVMELMTFGAKPYDGIPAREIPDLLEKGERLPPOPICTIDVYMI 954
QY 961 MYKCMWIDSECRPRRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MYKCMWIDSECRPRRELVSERMRARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDM 1014
QY 1021 GDLVDAEEYLVPOQGFCDPAPAGCMVHRRSRSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEEYLVPOQGFCDPAPAGCMVHRRSRSSSTRSGGDLTLGLEPSEEEAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPTVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGLQSLPHTHPSPLQRYSEDPTVPLPSETDGYVAPLTC 1134
QY 1141 POPEYVNOQDVRPPOPPSPREGPLPAARPAAGATILERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNOQDVRPPOPPSPREGPLPAARPAAGATILERAKTLSPGKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPOGGGAPOPHPPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGGAPOPHPPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
Db 1255 V 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human: HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stomal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX

PI Erickson S, Schwall R, Sliwowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

SQ Sequence 1255 AA;

Query Match 98.1%; Score 6708; DB 23; Length 1255;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 1242; Conservative 2; Mismatches 11; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLRPASPETHLDMRLHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQLRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLTORNPOLCYQDTILWKDIPKKNQLA 180
Db 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLTORNPOLCYQDTILWKDIPKKNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGACRCKPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGACRCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVYNTDTFESMPNPE-----GRYTFGAS 294
QY 301 CVTACPNYVLTSDVGSCTLVCPPLHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYVLTSDVGSCTLVCPPLHNOEVTAEQGTORCEKSPCARVCYGLGMEHLREVR 354
QY 361 AVTSANIOEFAGCKKIFGSLAFPLPSFDGDPASNTAPLOPEQLQVFETLEEITGYLYISA 420
Db 355 AVTSANIOEFAGCKKIFGSLAFPLPSFDGDPASNTAPLOPEQLQVFETLEEITGYLYISA 414
QY 421 WPSLPLDSVFNQLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSSGLALIHNTHL 480
Db 415 WPSLPLDSVFNQLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSSGLALIHNTHL 474
QY 481 CFVHTVPHWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQF 540
Db 475 CFVHTVPHWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQF 534
QY 541 LRQGEVCEECRVLOGLPREYVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPP 600
Db 535 LRQGEVCEECRVLOGLPREYVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPP 594
QY 601 FCVARCPGSKVPDLSYMPKFWPDEBEGACQPCPINCTHSCVLDLDDKGCAPASORASPLTSI 660

Db	595	FCVACPCSVKPOLSYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGCQAEQASPLTSI	
Qy	661	VSANVGIILLVWLVGVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI	
Db	655	VSANVGIILLVWLVGVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRI	
Qy	721	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANKETLDEAYVM	
Db	715	LKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANKETLDEAYVM	
Qy	781	AGVGSYVSRLLGICLTSTVQLVTQMPYCGCLLDHVRENRGRIGSODLLNWCWQIAKMS	
Db	775	AGVGSYVSRLLGICLTSTVQLVTQMPYCGCLLDHVRENRGRIGSODLLNWCWQIAKMS	
Qy	841	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESI	
Db	835	YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKMMALESI	
Qy	901	LRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMI	
Db	895	LRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMI	
Qy	961	MVKCMWIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDM	
Db	955	MVKCMWIDSECRPRFRELVSFERSMARDPQRFVVIQNEIDLGASPLDSTFYRSLLDDDM	
Qy	1021	GLVDAAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP	
Db	1015	GLVDAAEYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSP	
Qy	1081	LAPSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTC	
Db	1075	LAPSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVPLPSETDGVVAPLTC	
Qy	1141	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVPFPGGAVENP	
Db	1135	POPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVPFPGGAVENP	
Qy	1201	EYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	
Db	1195	EYLTPOGGAAPQHPHPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP	
Qy	1261	V 1261	
Db	1255	V 1255	
RESULT 5			
AAW01111			
ID	AAW01111	standard; Protein; 1255 AA.	
XX	AC	AAW01111;	
XX	DT	01-JAN-1997 (first entry)	
XX	DE	HER-2/new protein.	
KW	KW	HER-2/new; c-erbB1; p185; oncogene; tyrosine protein kinase; breast cancer; ovary cancer; colon cancer; lung cancer; prostate cancer; immunisation; tumour; vaccine; vector.	
OS	OS	Homo sapiens.	
FT	FT	Key	Location/Qualifiers
FT	FT	Domain	676..1255
FT	FT		/label= Intracellular domain
XX	XX		/note= "claimed domain, useful for immunisation"
PN	PN	WO9630514-A1.	
PD	PD	03-OCT-1996.	
XX	XX		

PF	28-MAR-1996;	96WO-USO1689.	
XX	31-MAR-1995;	95US-0414417.	
XX	(UNI*) UNIV WASHINGTON.		
XX	Cheever MA, Dieis ML;		
XX	WPI: 1996-455361/45.		
DR	N-PSDB; AAT40739.		
XX	DNA encoding HER-2-neu poly:peptide(s) - used for prevention or treatment of malignancies with which the HER-2/neu oncogene is associated		
PS	Claim 2; Page 56-61; 71pp; English.		
XX	Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is the product of the HER-2/neu oncogene (see also AAT40739). The protein is over-expressed in various cancers, including breast, ovarian, colon, lung and prostate. The intracellular domain of the protein can be used to immunise an animal against a malignancy with which the oncogene is associated. The polypeptide can be produced in transformed host cells for use in immunisation. Alternatively, animal cells are transfected in vivo or ex vivo with a viral vector that directs expression of the polypeptide.		
XX	Sequence 1255 AA;		
Qy	Query Match	98.0%; Score 6702; DB 17; Length 1255;	
Db	Best Local Similarity	98.3%; Pred. No. 0;	
Qy	Matches 1240; Conservative	3; Mismatches 12; Indels 6; Gaps 1;	
1	MELAAALCRWGLLLALLPGCAASTQVCTGDMKRLRIPASPTHLDMLRHLYQGCQVVGQNL	60	
1	MELAAALCRWGLLLALLPGCAASTQVCTGDMKRLRIPASPTHLDMLRHLYQGCQVVGQNL	60	
61	ELTYLPTNASLSFLODIOEVQGVYLI AHNVQVPLQRLRIRVGTOLFDNALAVLDNG	120	
61	ELTYLPTNASLSFLODIOEVQGVYLI AHNVQVPLQRLRIRVGTOLFDNALAVLDNG	120	
121	DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQRNPQLCYQDTILWKDIFHKQNQLA	180	
121	DPLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIQRNPQLCYQDTILWKDIFHKQNQLA	180	
181	LTLIDTNRSRACHPCSPMKSGRCRSGESSEDQSLTRTVACGACRCKGPLPTDCHEQC	240	
181	LTLIDTNRSRACHPCSPMKSGRCRSGESSEDQSLTRTVACGACRCKGPLPTDCHEQC	240	
241	AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFMLRVPKVSASHLERYTFGAS	300	
241	AAGCTGPKHSDCLACLFHNSGICELHCPALVFNFTVSFMLRVPKVSASHLERYTFGAS	294	
301	CVTACPNYLSLDVSGSCTLVCPFLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR	360	
295	CVTACPNYLSLDVSGSCTLVCPFLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVR	354	
361	AVTSANIQEFACKKIFGSLAPLPSFDGDPASNTAPLOPQLOVPETLEEITGYLYISA	420	
355	AVTSANIQEFACKKIFGSLAPLPSFDGDPASNTAPLOPQLOVPETLEEITGYLYISA	414	
421	WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRGLSGGLAIHHNTHL	480	
415	WPDLSPLDSVFONLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRGLSGGLAIHHNTHL	474	
481	CFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLAGCHQLCARGHCGWPGPTQCVNCSQF	540	
475	CFVHTVPMDQLFRNPHQALLHTANRPEDECVGEGLAGCHQLCARGHCGWPGPTQCVNCSQF	534	
541	LRGQECVBECEVLOGLPREYVNAHCLPCHPCQFONGSVTCFPGEADQCVACAHYKDP	600	
535	LRGQECVBECEVLOGLPREYVNAHCLPCHPCQFONGSVTCFPGEADQCVACAHYKDP	594	

Qy 601 FCVARCPSGVKPDLSYMPDIKFPDEBEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTISI 660
Db 595 FCVARCPSGVKPDLSYMPDIKFPDEBEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTISI 654
Qy 661 VSAVVGILLVVVLGVVGVGILLIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAOQRI 720
Db 655 ISAVVGILLVVVLGVVGVGILLIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAOQRI 714
Qy 721 LKETELRKVKVLSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLSGAFGTGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVM 774
Qy 781 AGVSPYVSRLLGLCLSTVOLVTLMPYGCGLLDHVRENRLGSGODLLNNCMQIAKMS 840
Db 775 AGVSPYVSRLLGLCLSTVOLVTLMPYGCGLLDHVRENRLGSGODLLNNCMQIAKMS 834
Qy 841 YLEOVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVKPKWMALES 900
Db 835 YLEOVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVKPKWMALES 894
Qy 901 LRRFRTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRFRTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGAAKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGAAKGLQSLPTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 PQEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKDVPAFGAVENP 1200
Db 1135 PQEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKDVPAFGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 6
AAW92406
AC AAW92406 standard; Protein; 1255 AA.
AC AAW92406;
DT 21-APR-1999 (first entry)
XX Human HER-2/neu oncogene protein.
DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
KW malignancy; treatment; tumour.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Region 676..1255
FT /note= "region which elicits immune response"
XX
PN US5869445-A.
XX
PD 09-FEB-1999.
XX
PF 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.
PR 17-MAR-1993; 93US-0033644.
PR 12-AUG-1993; 93US-0106112.
PR 31-MAR-1995; 95US-0414417.
XX (UNIM) UNIV WASHINGTON.
PA Cheever MA, Disis ML;
XX MPI; 1999-152835/13.
XX N-PSDB; AAX01912.
PT Use of HER-2/neu polypeptides - for eliciting an immune response to
PT an HER-2/neu associated malignancy, particularly for treating or
PT preventing tumours
XX Claim 3; Column 31-38; 26pp; English.
XX This sequence represents the human HER-2/neu oncogene protein. A fragment
CC of this protein is used in a method for eliciting or enhancing an immune
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 20; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPPGAASCTCTGDMKRLPASPTHLMRLHYGQCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTCTGDMKRLPASPTHLMRLHYGQCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVLNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVAGGACARCKGLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVAGGACARCKGLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNNFTVSWLRVPKVSASHLEYRTFGAS 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPE-----GRYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFTLEITGYLYISA 420
Db 355 AVTSANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFTLEITGYLYISA 414
Qy 421 WPDLSPLDSLVFQNLQVIRGRILHNGAYSLSLQGLGISWGLRSURELSGLALIHNNTHL 480
Db 415 WPDLSPLDSLVFQNLQVIRGRILHNGAYSLSLQGLGISWGLRSURELSGLALIHNNTHL 474
Qy 481 CFVHTVPWDLFRNPHQALLHTANRPEDECVGEGACHQLCARGCHGPGPTQCVNCSQF 540
Db 475 CFVHTVPWDLFRNPHQALLHTANRPEDECVGEGACHQLCARGCHGPGPTQCVNCSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTFGEADQCVACAHYKDDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCTFGEADQCVACAHYKDDP 594

Qy 601 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 660
Db 595 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNOAQMRI 720
Db 655 ISAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLRLLQETELVEPLTPSGAMPNOAQMRI 714
Qy 721 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYVM 774
Qy 781 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLLDHVRNRLGSGQDLNMCQIAKMS 840
Db 775 AGVSPYVSRLLGLCLTSTVOLVQLMPYGCGLLDHVRNRLGSGQDLNMCQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNLVKSPNVKVTDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEVDRLVHRDLAARNLVKSPNVKVTDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMIDSECRPRFRELVSSEFSRMARDPQRVVIQNEDELGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMIDSECRPRFRELVSSEFSRMARDPQRVVIQNEDELGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCDPAPAGAGVHHRHSSSTRSGGDLITLGLPSEBEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCDPAPAGAGVHHRHSSSTRSGGDLITLGLPSEBEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPQLORYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPQLORYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAGGAVENP 1200
Db 1135 PQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSFAFNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 7
ID AAB21198
XX AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
DE
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS Homo sapiens.
XX
PN WO200004899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX

(CORI-) CORIXA CORP.
PA (SMIK) SMITHLINE BEECHAM.
XX
PI Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6702; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRASPETHLDMRLHYQCCVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRASPETHLDMRLHYQCCVVOGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVQYVLIHNVQVPLQRLRVIRGTQFEDNVALAVLNG 120
Db 61 ELTYLPTNASLFLQDIQEVQYVLIHNVQVPLQRLRVIRGTQFEDNVALAVLNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGGCARCKGPLTDCCHQOC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVAGGCARCKGPLTDCCHQOC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSPWLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSPWLRVPKVSASHLERYTFGAS 294
Qy 301 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPOLQVFTLEITGYLYISA 420
Db 355 AVTSANTQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPOLQVFTLEITGYLYISA 414
Qy 421 WPDSLPLDSVFQNLQVIRGRILHNGAYSLSLQGLGISWLGSLRSLRELGGSLALHNNTHL 480
Db 415 WPDSLPLDSVFQNLQVIRGRILHNGAYSLSLQGLGISWLGSLRSLRELGGSLALHNNTHL 474
Qy 481 CFVHTVPMDQLFRNPQALLHTANRPBECVGEGLACHQLCARGHGWPGTQCVCNCSQF 540
Db 475 CFVHTVPMDQLFRNPQALLHTANRPBECVGEGLACHQLCARGHGWPGTQCVCNCSQF 534
Qy 541 LRQGECEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db 535 LRQGECEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVACPSGVKPDLSYMPKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI 660

Db 595 FCVARCPGKVPDLSTYMP1WKFPEEGACOPCPINCTHSCVDLDDKGPQASPLTSI 654
Qy 661 VSAVVGILLVVLGVWFGLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNQAWRI 720
Db 655 ISAVVGILLVVLGVWFGLIKRQOKIRKTYTMRLLQETELVEPLTPSGAMPNQAWRI 714
Qy 721 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVM 774
Qy 781 AGVSPYVRLGLICLTSTVQLTQMLPYGCLLDHVRENRLGSDQLLNWCMQIAKMS 840
Db 775 AGVSPYVRLGLICLTSTVQLTQMLPYGCLLDHVRENRLGSDQLLNWCMQIAKMS 834
Qy 841 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMWALESI 900
Db 835 YLEDRVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMWALESI 894
Qy 901 LRRRFTHOSDWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHOSDWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDM 1014
Qy 1021 GDLVDAEYLVPQGGFCPPDAPACAGGMVHRHSSSTRSGGDLTLGLPSEEEARSP 1080
Db 1015 GDLVDAEYLVPQGGFCPPDAPACAGGMVHRHSSSTRSGGDLTLGLPSEEEARSP 1074
Qy 1081 LAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPQRYSEDPTVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPQRYSEDPTVPLPSETDGYVAPLTC 1134
Qy 1141 POEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDDVP 1260
Db 1195 EYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGDDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 8
AAV84780
ID AAV84780 standard; Protein; 1255 AA.
XX
AC AAV84780;
XX
XX 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
XX wound healing.
XX
OS Homo sapiens.
XX
XX WO200020579-A1.
XX
XX 13-APR-2000.
XX
XX 01-OCT-1999; 99WO-CA00912.
XX
XX 02-OCT-1998; 98US-0165192.
XX

PA (UYMC-) UNIV MCMASTER.
XX Muller WJ, Siegel PM;
XX WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbb-2 receptor protein. The
protein has an in-frame deletion of 16 amino acids, 2 of which are
conserved cysteine residues, compared to the unspliced protein. The
erbb-2 polynucleotide is used to construct probes for detecting
disorders of cell transformation such as cancer. Antibodies to the
protein may be used to detect SPLICE erbb-2 in a sample. Agents
(e.g. antisense oligonucleotides) which inhibit the expression of
SPLICE erbb-2 are useful for reducing tumor cell proliferation and
treating cancer. Substances which stimulate SPLICE erbb-2 are useful
for treating conditions involving damaged cells including conditions
in which degeneration of tissue occurs, such as arthropathy, bone
resorption, inflammatory diseases, degenerative disorders of the
central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 21; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYGQCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVRQVPLQRLRIVRGTLQFEDNVYALVLDNG 120
Db 61 ELTYLPTNASLFLQDIEQVQYVLIHNOVRQVPLQRLRIVRGTLQFEDNVYALVLDNG 120
Qy 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNOLA 180
Db 121 DPLNTPVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNOLA 180
Qy 181 LTLDITNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Db 181 LTLDITNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVFNNFTVSFVLRVPKVSASHLERYTCAS 300
Db 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFFSMPNPE-----GRYTFGAS 294
Qy 301 CVTACPNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTIQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFETLEBITGYLYISA 420
Db 355 AVTSANTIQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFETLEBITGYLYISA 414
Qy 421 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSRELGSGLALIHNNTHL 480
Db 415 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGSLRSRELGSGLALIHNNTHL 474
Qy 481 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVCNCSQF 540
Db 475 CFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVCNCSQF 534
Qy 541 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVCACAHYKDDP 600
Db 535 LRGOECVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVCACAHYKDDP 594
Qy 601 FCVARCPGKVPDLSTYMP1WKFPEEGACQPCPINCTHSCVDLDDKGPQASPLTSI 660

Db	595	FCVARCPSGVKPDLSYPMKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSI	654
Qy	661	VSAVGILLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRI	720
Db	655	ISAVVGILLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRI	714
Qy	721	LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEIIDEAVYM	780
Db	715	LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEIIDEAVYM	774
Qy	781	AGVSPYVRLGLICLTSTVOLTPMYPYGCLLDHRVRENRLGSDQLLNKCMQIAKMS	840
Db	775	AGVSPYVRLGLICLTSTVOLTPMYPYGCLLDHRVRENRLGSDQLLNKCMQIAKMS	834
Qy	841	YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVKPIKMALES	900
Db	835	YLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVKPIKMALES	894
Qy	901	LRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI	960
Db	895	LRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI	954
Qy	961	MVKCWMIDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDM	1020
Db	955	MVKCWMIDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDM	1014
Qy	1021	GDLDVAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSCGGDLTLGLEPSEEAAPRS	1080
Db	1015	GDLDVAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSCGGDLTLGLEPSEEAAPRS	1074
Qy	1081	LAPSEGAGSDVFDGDLGGAAGKQLSPLTHDPSPQLQYSEDPVPLPSETDGYVAPLTC	1140
Db	1075	LAPSEGAGSDVFDGDLGGAAGKQLSPLTHDPSPQLQYSEDPVPLPSETDGYVAPLTC	1134
Qy	1141	POPEYVNOPVVRPOPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVPFAGAVENP	1200
Db	1135	POPEYVNOPVVRPOPSPREGPLPAARPAGATLERAKTSLSPGKNGVVKDVPFAGAVENP	1194
Qy	1201	EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYVGLDVP	1260
Db	1195	EYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYVGLDVP	1254
Qy	1261	V	1261
Db	1255	V	1255
RESULT 9			
ID	AA885458	standard; Protein; 1255 AA.	
AC	AA885458;		
XX	25-SEP-2001	(first entry)	
DE	Human HER-2/neu protein.		
KW	Antigen-presenting cell; immunogenic; immune response; HER-2/neu; oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.		
OS	Homo sapiens.		
FN	WO200153463-A2.		
PD	26-JUL-2001.		
XX	19-JAN-2001; 2001WO-US01850.		
XX	21-JAN-2000; 2000US-0177545.		
PA	(CORI-) CORIYA CORP.		
XX			

PI	Cheever MA, Hand-Zimmermann S;	
XX	WPI; 2001-476112/51.	
DR	N-PSDB; AAH23392.	
XX	New antigen-presenting cells, useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly useful for treating or preventing cancer, e.g. breast cancer	
PT	Claim 2; Page 41-46; 49pp; English.	
XX	The invention provides an isolated antigen-presenting cell, which expresses at least an immunogenic portion of a polypeptide that produces an immune response to HER-2/neu protein. The antigen-presenting cells are useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly in treating or preventing malignancies in which the HER-2/neu oncogene is associated. Specifically, these are useful for treating or preventing cancer, e.g. breast cancer, ovarian, colon, lung or prostate cancers. The present sequence represents the human HER-2/neu protein (also known as p185 or c-erbB2).	
CC	Sequence 1255 AA;	
QY	Query Match 98.0%; Score 6702; DB 22; Length 1255;	
DB	Best Local Similarity 98.3%; Pred. No. 0;	
XX	Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;	
QY	1	MELAAALCRWGLLLALLPPGAASSTQVCTGTDMLKRLPASPTHLDMLRLHYGCGVVOGNL 60
DB	1	MELAAALCRWGLLLALLPPGAASSTQVCTGTDMLKRLPASPTHLDMLRLHYGCGVVOGNL 60
QY	61	ELTYLPTNASLFLQDIQEVGYVLIHANOVQVPLQRLRVIRGTOLFEDNYALAVLDNG 120
DB	61	ELTYLPTNASLFLQDIQEVGYVLIHANOVQVPLQRLRVIRGTOLFEDNYALAVLDNG 120
QY	121	DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNQLA 180
DB	121	DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLTORNPOLCYQDTILWKDIFHKNNQLA 180
QY	181	LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHQOC 240
DB	181	LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTDCCHQOC 240
QY	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVFNFTVTSFWLVRVPKVSASHLERYTFGAS 300
DB	241	AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPE-----GRYTFGAS 294
QY	301	CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
DB	295	CVTACPNYLSLTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
QY	361	AVTSANIQEFAGCKKIFGSLAFLESDFGDPPASNTAPLOPQLOVPETLEETIGYLYISA 420
DB	355	AVTSANIQEFAGCKKIFGSLAFLESDFGDPPASNTAPLOPQLOVPETLEETIGYLYISA 414
QY	421	WPDLSLPDLVSFQNLQVIRGRILHNGAYSLTQGLGISWLGSLRSLRELGSGLAIHNTHL 480
DB	415	WPDLSLPDLVSFQNLQVIRGRILHNGAYSLTQGLGISWLGSLRSLRELGSGLAIHNTHL 474
QY	481	CFVHTVPWDQLFRNPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPGPTQVCNCSQF 540
DB	475	CFVHTVPWDQLFRNPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPGPTQVCNCSQF 534
QY	541	LRGQECYBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 600
DB	535	LRGQECYBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDDP 594
QY	601	FCVARCPSGVKPDLSYMPKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSI 660
DB	595	FCVARCPSGVKPDLSYMPKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSI 654
QY	661	VSAVGILLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQMRI 720
XX	:	


```
Db 655 ISAVVGVLLVVVLGVVFGILLIKRQKIRKYMRRLLQETELVEPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEAYVM 780
Db 715 LKETELRKVKVLSGAFGTGVYGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEAYVM 774
Qy 781 AGVCSPVVSRLLGICLTSTVOLVTLMPYGCGLLDHVRENRGRISQDLNWKQIAKMS 840
Db 775 AGVCSPVVSRLLGICLTSTVOLVTLMPYGCGLLDHVRENRGRISQDLNWKQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNLVKSPNHNKIDTFGLARLLDIDETEHADGGKVPKMWALESI 900
Db 835 YLEVDRLVHRDLAARNLVKSPNHNKIDTFGLARLLDIDETEHADGGKVPKMWALESI 894
Qy 901 LRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1020
Db 955 MVKCMWIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDM 1014
Qy 1021 GDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFPCPDPAFCAGGMVHRRHSSSTRSGGDLTLGLEPSEEEAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSFLQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEGAGSDVFDGDLGMAAGKQSLPHTDPSFLQRYSEDPTVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNOPDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVDVPAFGAVENP 1200
Db 1135 POPEYVNOPDVRPOPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVDVPAFGAVENP 1194
Qy 1201 EYLTPQGAAPQPPHPPAFPAFONLYWDDPPERGAPPSTFKGTPTTAENPEYVLGLDVP 1260
Db 1195 EYLTPQGAAPQPPHPPAFPAFONLYWDDPPERGAPPSTFKGTPTTAENPEYVLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein: 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN W0200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI; 2001-374995/39.
```

```
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6702; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
```

```
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60
Qy 61 ELYLPTNASLSFLQDIEQVGVYLIAHNOVRQVPLORLIRVRGTQLPEDNVALLVDNG 120
Db 61 ELYLPTNASLSFLQDIEQVGVYLIAHNOVRQVPLORLIRVRGTQLPEDNVALLVDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSPACHPCSPMKYSGRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSPACHPCSPMKYSGRCWGESSEDCQSLTRTVCGAGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVFNNFTVSFWLRVPKVSASHLERTFCAS 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVFNNFTVSFWLRVPKVSASHLERTFCAS 300
Qy 301 CVTACPNYLLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEFAGCKKIFGSLAFIPESPDGDPASNTAPLQPEQLQVFPETLEITGYLYISA 420
Db 355 AVTSANIQEFAGCKKIFGSLAFIPESPDGDPASNTAPLQPEQLQVFPETLEITGYLYISA 414
Qy 421 WPDLSPLDSVFNQLQVIRGRILHNGAYSLTLQGISWLGSLRLSRLGSLALHNNTHL 480
Db 415 WPDLSPLDSVFNQLQVIRGRILHNGAYSLTLQGISWLGSLRLSRLGSLALHNNTHL 474
Qy 481 CFVHTVPWDLFRNPHQALLHTANRPEDECVGELACHQLCARGHCGMGPGPTQCVCNSQF 540
```

Db 475 CFVHTVPMWDLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRQGECEBECVRLQGLPREYNARHCLCHPECPQNGSVTCFGEADQCACAHYKDDP 600
Db 535 LRQGECEBECVRLQGLPREYNARHCLCHPECPQNGSVTCFGEADQCACAHYKDDP 594
Qy 601 FCVARCPGKPDLSYMPIMKFPDDEGACQPCINCHSCVDLDDKGCPCABQASPLTSI 660
Db 595 FCVARCPGKPDLSYMPIMKFPDDEGACQPCINCHSCVDLDDKGCPCABQASPLTSI 654
Qy 661 VSAVVGILLVVVLGVVFGILLKRRQOKIRKKTMRLLQETELVEPLTSPGAMPNQAQWRI 720
Db 655 ISAVVGILLVVVLGVVFGILLKRRQOKIRKKTMRLLQETELVEPLTSPGAMPNQAQWRI 714
Qy 721 LKETELRKVKVLSGAFVTVKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLSGAFVTVKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 774
Qy 781 AGVSPYVRLIGILCTSTVOLVTLQMPYGCLLDHVRENRLGSGQDLLNMCQIAKMS 840
Db 775 AGVSPYVRLIGILCTSTVOLVTLQMPYGCLLDHVRENRLGSGQDLLNMCQIAKMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALESI 894
Qy 901 LRRRFTHQSVDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db 895 LRRRFTHQSVDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Qy 961 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDD 1020
Db 955 MVKCMIDSECRPRELVSFSEMRARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDD 1014
Qy 1021 GDLVDAEYLYPQOGFFCPDPAGGAGGVHRRSSSTRSCGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLVDAEYLYPQOGFFCPDPAGGAGGVHRRSSSTRSCGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGDLGAAKGLQSLPHTPSPQLRYSEDPVPLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGDLGAAKGLQSLPHTPSPQLRYSEDPVPLPSETDGYVAPLTC 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFAGGAVENP 1194
Qy 1201 EYLTPOGGAAPQHPHPPAFSAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db 1195 EYLTPOGGAAPQHPHPPAFSAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
OS Homo sapiens.
XX
XX W0200222636-A1.
PN
XX

PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseert LM;
XX
DR WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targeted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Qy 61 ELTYLPTNASLFIQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLFIQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSGRCWGESSEDCQSLTRTVACAGGCARCKGPLTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSGRCWGESSEDCQSLTRTVACAGGCARCKGPLTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNNFTVSPFMRVPKVSASHLRYTFGAS 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNNFTVSPFMRVPKVSASHLRYTFGAS 294
Qy 301 CVTACPNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSTDVGSCTLVCPHLNQVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANTIOEPAGCKIFGSLAFIPESFDGPPASNTAPLOPQLOVFTLEBITGYLYISA 420
Db 355 AVTSANTIOEPAGCKIFGSLAFIPESFDGPPASNTAPLOPQLOVFTLEBITGYLYISA 414
Qy 421 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 480
Db 415 WPDSLPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLGSLALIHNNTHL 474
Qy 481 CFVHTVPMWDLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 540
Db 475 CFVHTVPMWDLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNSQF 534
Qy 541 LRQGECEBECVRLQGLPREYNARHCLCHPECPQNGSVTCFGEADQCACAHYKDDP 600

Db 535 LRQGECECKVLQGLPREYNARHCUPCHPECPQNGSVTCFGEADQCACARHYKDP 594
Qy 601 FCVACRCPGKPDLSYMPINWKPDEEGACQCPINCTHSCVDLDDKGCAPQASPLTISI 660
Db 595 FCVACRCPGKPDLSYMPINWKPDEEGACQCPINCTHSCVDLDDKGCAPQASPLTISI 654
Qy 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOQMRI 720
Db 655 ISAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVPLTPSGAMPNOQMRI 714
Qy 721 LKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAVYM 780
Db 715 LKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEIIDEAVYM 774
Qy 781 AGVSPVYRLLGLCLTSTVOLVQLMPYGCLLDHRNRCRLGSDQLLNWCMQIAKMS 840
Db 775 AGVSPVYRLLGLCLTSTVOLVQLMPYGCLLDHRNRCRLGSDQLLNWCMQIAKMS 834
Qy 841 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESI 900
Db 835 YLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESI 894
Qy 901 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 960
Db 895 LRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMI 954
Qy 961 MVKCMWIDSECRPRELVSEFSRMDPQRFVVIQNEGLGSPASPLDSTFYRSILLEDDM 1020
Db 955 MVKCMWIDSECRPRELVSEFSRMDPQRFVVIQNEGLGSPASPLDSTFYRSILLEDDM 1014
Qy 1021 GDLVDAEYLVPOQGFCDPAPAGAGWVHRHRSSTRSGGDLITLGLPSEEEAPRSP 1080
Db 1015 GDLVDAEYLVPOQGFCDPAPAGAGWVHRHRSSTRSGGDLITLGLPSEEEAPRSP 1074
Qy 1081 LAPSEAGSDVFDGDLGWAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCS 1140
Db 1075 LAPSEAGSDVFDGDLGWAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCS 1134
Qy 1141 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPNGKGVVVDVFAFGAVENP 1200
Db 1135 POPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTLPNGKGVVVDVFAFGAVENP 1194
Qy 1201 EYLTPQGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSPFTKGTPTASNPEYLGLDVP 1260
Db 1195 EYLTPQGAAPQHPHPPAFSPAFNLVYWDODPPERGAPSPFTKGTPTASNPEYLGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
DT 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
DE Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
KW Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX

PD 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
PF 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-234428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedwick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
PT Disclosure; Page 114-117; 129pp; English.
PS The invention relates to an isolated Her-2/Neu polypeptide composition
XX effective for eliciting an immune response. The invention is useful for
XX eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
XX present sequence is human Her-2/neu protein.
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPTHDLMLRHLVQGCQVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMLRLPASPTHDLMLRHLVQGCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGVYLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGVYLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 180
Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSWLRVPKVSASHLRYTTCAS 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSWLRVPKVSASHLRYTTCAS 294
Qy 301 CVTACPNYLSLTDVSGCTTLCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLSLTDVSGCTTLCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIQEPAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEITGELYISA 420
|||||

QY 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
DB 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCGARCKGPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCGARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSWLVRVPKVSASHLEYTFGAS 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVFNFTVSWLVRVPKVSASHLEYTFGAS 300
QY 301 CVTACPNYLSTDVGSCTLCVPLHNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 360
DB 295 CVTACPNYLSTDVGSCTLCVPLHNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVR 354
QY 361 AVTSANTQEFAGCKKIFGSLAFLPESFDGPPASNTAPLOPEQLQVFTLEETIGLYISA 420
DB 355 AVTSANTQEFAGCKKIFGSLAFLPESFDGPPASNTAPLOPEQLQVFTLEETIGLYISA 414
QY 421 WPOSPLDSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHL 480
DB 415 WPOSPLDSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLGSLALIHNTHL 474
QY 481 CFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGELACHQLCARGHCWGPPTQCVNCSQF 540
DB 475 CFVHTVPWDOLFRNPHQALLHTANRPEDECVGEGELACHQLCARGHCWGPPTQCVNCSQF 534
QY 541 LRQOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 600
DB 535 LRQOECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDDP 594
QY 601 FCVAPCSGVKPLSYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKCPAEQASPLTSI 660
DB 595 FCVAPCSGVKPLSYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKCPAEQASPLTSI 654
QY 661 VSAVVGILLVVLGVVFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGMNPQOMRI 720
DB 655 ISAVVGILLVVLGVVFGIILKRRQKIRKYTWRRLLQETELVEPLTPSGMNPQOMRI 714
QY 721 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 780
DB 715 LKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVM 774
QY 781 AGVSPVSVRLGICLTSTVQLVTLQMPYGCCLLDHVRENRGRGSLQDLNWCWQIAKMS 840
DB 775 AGVSPVSVRLGICLTSTVQLVTLQMPYGCCLLDHVRENRGRGSLQDLNWCWQIAKMS 834
QY 841 YLEDVRLVHRDLAARNVLSKPNHVKITDFGLARLLDIDETEHADGSKVPKIKWMALESI 900
DB 835 YLEDVRLVHRDLAARNVLSKPNHVKITDFGLARLLDIDETEHADGSKVPKIKWMALESI 894
QY 901 LRRFRTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 960
DB 895 LRRFRTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMI 954
QY 961 MVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEM 1020
DB 955 MVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEM 1014
QY 1021 GOLVDAEEYLVPOQGFCDPAPACAGMWHHRSSSTRSGGDLTLGLPEPSEEAAPRSP 1080
DB 1015 GOLVDAEEYLVPOQGFCDPAPACAGMWHHRSSSTRSGGDLTLGLPEPSEEAAPRSP 1074
QY 1081 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPFLQRYSEDPTVPLPSETDGVVAPLTC 1140
DB 1075 LAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPFLQRYSEDPTVPLPSETDGVVAPLTC 1134
QY 1141 POEYVYNQDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAGGAVENP 1200

DB 1135 POEYVYNQDVRPQPPSPREGPLPAARPAGATLERPKTILSPGKNGVVKDVFAGGAVENP 1194
QY 1201 EYLTPQCGAAPQHPPPAFSPAFDNLVYWOODPBERGAPPSTFKGTPTAENPEYLGLDVP 1260
DB 1195 EYLTPQCGAAPQHPPPAFSPAFDNLVYWOODPBERGAPPSTFKGTPTAENPEYLGLDVP 1254
QY 1261 V 1261
DB 1255 V 1255
RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AAU77114;
XX 05-JUN-2002 (first entry)
DT Human Her-2/neu polypeptide.
DE Human Her-2/neu polypeptide.
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX Homo sapiens.
OS
XX
PN WO200213847-A2.
XX 21-FEB-2002.
PD 13-AUG-2001; 2001WO-US25408.
XX 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
DR N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 98.0%; Score 6702; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1240; Conservative 3; Mismatches 12; Indels 6; Gaps 1;
QY 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLVQGCQVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASTQVCTGDMKRLPASPETHLDMLRHLVQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGSESDCQSLTRTVACAGGCARCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGSESDCQSLTRTVACAGGCARCKGPLPTDCCHQC 240
Qy 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVFNNFTVS FMLRVPKVSASHLERYTFGAS 300
Db 241 AAGCTGPRKSDCLACLFHNSGICELHCPALVTYNTDTFESMPNE-----GRYTFGAS 294
Qy 301 CVTACPNYLLSTDVSGSLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
Db 295 CVTACPNYLLSTDVSGSLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 354
Qy 361 AVTSANIOEFAGCKKI FGSFLAFLESFDGDPASNTAPLOEQLOVFETLBEITGYLYTISA 420
Db 355 AVTSANIOEFAGCKKI FGSFLAFLESFDGDPASNTAPLOEQLOVFETLBEITGYLYTISA 414
Qy 421 WPDLSPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLRELSGLALIHNNTHL 480
Db 415 WPDLSPLDSVFQNLQVIRGRILHNGAYSLLTQGLGISWLGRLSRLRELSGLALIHNNTHL 474
Qy 481 CFVHTVPMDOLFNRPHQALLHTANRPEDECVGEGGLACHQLCARHCWGPGTQCVCNCSQF 540
Db 475 CFVHTVPMDOLFNRPHQALLHTANRPEDECVGEGGLACHQLCARHCWGPGTQCVCNCSQF 534
Qy 541 LRQGCVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDKPP 600
Db 535 LRQGCVEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDKPP 594
Qy 601 FCVARCPGKVPDLSYMPIKFPDDEGACQPCINCTHSCVDLDDKGCPCAEQASPLTSI 660
Db 595 FCVARCPGKVPDLSYMPIKFPDDEGACQPCINCTHSCVDLDDKGCPCAEQASPLTSI 654
Qy 661 VSAVGILLVVLGVVVGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRI 720
Db 655 ISAVVGILLVVLGVVVGILIKRROOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRI 714
Qy 721 LKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 780
Db 715 LKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVM 774
Qy 781 AGVGSPPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRRGLSGQDLLNWCMIQAKGMS 840
Db 775 AGVGSPPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRRGLSGQDLLNWCMIQAKGMS 834
Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGGKVPKKNWALESI 900
Db 835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHYHADGGKVPKKNWALESI 894
Qy 901 LRRRTHOSDVMYSGVTWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICITIDVYMI 960
Db 895 LRRRTHOSDVMYSGVTWELMTFGAKPYDGIIPAREIPDLLEKGBRLPOPPICITIDVYMI 954
Qy 961 MVKCMIDSECRPRELVSFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDM 1020
Db 955 MVKCMIDSECRPRELVSFMRARDPQRFVVIQNEDLGPASPLDSTFYRSLEDDDM 1014
Qy 1021 GDLDVAEYLVPPQGFPCPDAPAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSP 1080
Db 1015 GDLDVAEYLVPPQGFPCPDAPAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSP 1074
Qy 1081 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDPSPQLQRYSEDPVLPSETDGYVAPLTC 1140
Db 1075 LAPSEGAGSDVFDGLGMAAGKLOSLPTHDPSPQLQRYSEDPVLPSETDGYVAPLTC 1134
Qy 1141 PQPEYVNPQDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVPFAFGAVENP 1200

Db 1135 PQPEYVNPQDVRPOPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVPFAFGAVENP 1194
Qy 1201 EYLTPQCGAAAPQHPHPPAFAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYILGLDVP 1260
Db 1195 EYLTPQCGAAAPQHPHPPAFAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYILGLDVP 1254
Qy 1261 V 1261
Db 1255 V 1255
RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX AAR39568;
AC AAR39568;
DT 07-FEB-1994 (first entry)
XX Sequence of c-erbB-2 tumour antigen.
DE Tumour antigen; c-erbB-2; glycoprotein.
KW Homo sapiens.
OS WO9316185-A.
PN 19-AUG-1993.
XX 05-FEB-1993; 93WO-US01055.
XX 06-FEB-1992; 92US-0831967.
XX (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX MPI, 1993-272889/34.
DR N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.
PT Disclosure; pages 48-54; 87pp; English.
XX c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents the location of a stop codon in AAQ46083.
XX Sequence 1433 AA;
SQ Query Match 97.4%; Score 6659; DB 14; Length 1433;
Best Local Similarity 97.8%; Pred No. 0;
Matches 1233; Conservative 5; Mismatches 17; Indels 6; Gaps 1;
Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRASPETHLDMRLHYQCVVQGNL 60
Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGCTDMKRLRASPETHLDMRLHYQCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGSESDCQSLTRTVACAGGCARCKGPLPTDCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGSESDCQSLTRTVACAGGCARCKGPLPTDCCHQC 240

Search completed: July 22, 2003, 08:41:06
Job time : 43.1589 secs

```
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVFNNFTVSFWLRVPKVSASHLERYTFGAS 300
Db      |||||
241 AAGCTGPKHSDCLACLFHNSGICELHCPALVNTYNTDTFESMPNPE-----GRYTFGAS 294
      |||||

Qy 301 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 360
Db      |||||
295 CVTACPNYLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVR 354
      |||||

Qy 361 AVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFFETLEITGYLYISA 420
Db      |||||
355 AVTSANTIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEHLQVFFETLEITGYLYISA 414
      |||||

Qy 421 WPSLPLDLSVFNQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHL 480
Db      |||||
415 WPSLPLDLSVFNQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLGSLALIHNNTHL 474
      |||||

Qy 481 CFVHTVPWDQLFRPHQALLHTANRPEDECVEGGLACHOLCARGHCWGPGPTQCVCNSQF 540
Db      |||||
475 SFVHTVPWDQLFRPHQALLHTANRPEDECVEGGLACHOLCARGHCWGPGPTQCVCNSQF 534
      |||||

Qy 541 LRQCECVEECRVLOGLPREYVNAHRLCPCHPECPQNGSVTCFGEADOCVACAHYKOPP 600
Db      |||||
535 LRQCECVEECRVLOGLPREYVNAHRLCPCHPECPQNGSVTCFGEADOCVACAHYKOPP 594
      |||||

Qy 601 FCVARCPGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSI 660
Db      |||||
595 FCVARCPGKVPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSI 654
      |||||

Qy 661 VSAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWRI 720
Db      |||||
655 ISAVVGILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWRI 714
      |||||

Qy 721 LKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVM 780
Db      |||||
715 LKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVM 774
      |||||

Qy 781 AGVSPYVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRCRLGSDLLNWCMIAGMS 840
Db      |||||
775 AGVSPYVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRCRLGSDLLNWCMIAGMS 834
      |||||

Qy 841 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKVPKWMMALESI 900
Db      |||||
835 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKVPKWMMALESI 894
      |||||

Qy 901 LRRRFTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 960
Db      |||||
895 LRRRFTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
      |||||

Qy 961 MVKCMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLEDDDM 1020
Db      |||||
955 MVKCMIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLEDDDM 1014
      |||||

Qy 1021 GDLVDAEYLVPOQGFPCPDAPCAGAGMVHHRHSSSTRSCGGDLTLGLPSEEEAPRSP 1080
Db      |||||
1015 GDLVDAEYLVPOQGFPCPDAPCAGAGMVHHRHSSSTRSCGGDLTLGLPSEEEAPRSP 1074
      |||||

Qy 1081 LAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1140
Db      |||||
1075 LAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCS 1134
      |||||

Qy 1141 POPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLPSPGNKVVDVFAFGAVENP 1200
Db      |||||
1135 POPEYVNPQDVRPQPPSPREGPLPAARPAGATLERPKTLPSPGNKVVDVFAFGAVENP 1194
      |||||

Qy 1201 EYLTPOCGAAPQHPHPPAFSPADNLYWDDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1260
Db      |||||
1195 EYLTPOCGAAPQHPHPPAFSPADNLYWDDPPPERGAPPSTFKGTPTAENPEYLGLDVP 1254
      |||||

Qy 1261 V 1261
Db      |
1255 V 1255
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-325-339-14
Perfect score: 6852
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKCTPTAENPEYLGLOVPV 1261
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6730	98.2	1255	1 A24571	protein-tyrosine k
2	5925	86.5	1260	1 TVRTNU	protein-tyrosine k
3	5916.5	86.3	1254	2 I48161	p-185 precursor
4	3150	46.0	1210	1 GQHUE	epidermal growth f
5	3123	45.6	1210	2 A53183	epidermal growth f
6	3099.5	45.2	1223	1 TVCHLV	epidermal growth f
7	2981.5	43.5	1308	2 A47233	epidermal growth f
8	2672	39.0	1166	1 S06142	epidermal growth f
9	2430.5	35.5	1342	2 A36223	protein-tyrosine k
10	2342.5	34.2	1339	2 JC4387	kinase-related tra
11	1766.5	25.8	698	1 TVFVLV	epidermal growth f
12	1703	24.9	604	1 TVYUHV	protein-tyrosine k
13	1647	24.0	544	2 S35745	protein-tyrosine k
14	1643.5	24.0	1330	1 QQFFE	protein-tyrosine k
15	1640	23.9	545	2 S00727	kinase-related tra
16	1623	23.7	540	2 B44776	protein-tyrosine k
17	1621	23.7	540	1 TVFVBB	protein-tyrosine k
18	1515	22.1	644	2 A36325	epidermal growth f
19	1300	19.0	1323	2 E88257	protein let-23 lim
20	1300	19.0	1374	2 S70712	protein-tyrosine k
21	1214	17.7	1369	2 S70713	protein-tyrosine k
22	1180	17.2	1717	1 A45558	epidermal growth f
23	1131	16.5	527	2 A42032	epidermal growth f
24	988.5	14.4	84	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.0	311	2 S13807	protein-tyrosine k
27	730	10.7	1363	2 T43220	protein-tyrosine k
28	698	10.2	1382	1 INHUR	insulin-like growt
29	693	10.1	1607	2 T43212	insulin receptor p

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein erb
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A>Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A>Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R:Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,
Science 230, 1132-1139, 1985
A>Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:G183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A>Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A>Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

insulin receptor p
insulin receptor p
protein-tyrosine k
insulin receptor-r
insulin receptor-r
insulin-like growt
insulin receptor -
insulin-like growt
insulin receptor (r
insulin receptor -
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
mouse developmenta
protein-tyrosine k

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL: NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
kinase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68.124.187.259.530.571.629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.2%; Score 6730; DB 1; Length 1255;
Best Local Similarity 98.4%; Pred. No. 1.6e-267;
Matches 1242; Conservative 8; Mismatches 4; Indels 8; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLPASPETHLDMRLHYGCGVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHNRQVPLQLRLVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLIHNRQVPLQLRLVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTCASPGGLRELQRLSITLKGVLIRNPOLCYODTILWKDIFHNKOLA 180
Db 121 DPLNNTPTVTCASPGGLRELQRLSITLKGVLIRNPOLCYODTILWKDIFHNKOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCGAGCARCKGPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCGAGCARCKGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMFNFTVSFWLR-V 359
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMFNH-----LREV 353

Qy 360 PKVSASHLEEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEEITGYLYIS 419
Db 360 PKVSASHLEEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEEITGYLYIS 413

Qy 420 AMPDLSPLSVFONLQVIRGRILHNGAYSITLQGLGTSWGLRSLRELGLSLALIHNTHT 479
Db 420 AMPDLSPLSVFONLQVIRGRILHNGAYSITLQGLGTSWGLRSLRELGLSLALIHNTHT 473

Qy 480 LCFVHTVPWDLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCGPQPTOCVNCQ 539
Db 480 LCFVHTVPWDLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCGPQPTOCVNCQ 533

Qy 540 FLRQECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPQADQCAVACHYKDP 599
Db 540 FLRQECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFPGPQADQCAVACHYKDP 593

Qy 600 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGPACQASPLTS 659
Db 600 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGPACQASPLTS 659

Db 594 PFCVARCPSPGKPDLSYMPITWKFDEGACQPCPINCTHSCVDLDDKGPACQASPLTS 653
Qy 660 IVSAVVGTLVWLVGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQR 719
Db 654 IISAVVGTLVWLVGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQR 713
Qy 720 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVAPIVAIKVLRNTPSKANKEILDEAYV 779
Db 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVAPIVAIKVLRNTPSKANKEILDEAYV 773
Qy 780 MAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVRNRRGLSGDQLLNWCMQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHVRNRRGLSGDQLLNWCMQIAKGM 833
Qy 840 SYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADCGKVPKRWMALES 899
Db 834 SYLEDVRLVHRDLAARNVLKSPNHVKTDFGLARLLDIDETEVHADCGKVPKRWMALES 893
Qy 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYM 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYM 953
Qy 960 IMVKCMWIDSECRPRPRELVSEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDD 1019
Db 954 IMVKCMWIDSECRPRPRELVSEFMRMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDD 1013
Qy 1020 MGDLDVAEEYLVPOGFCFDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRS 1079
Db 1014 MGDLDVAEEYLVPOGFCFDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRS 1073
Qy 1080 PLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOYSRSDPTVPLPSETDGYVAPLTC 1139
Db 1074 PLAPSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOYSRSDPTVPLPSETDGYVAPLTC 1133
Qy 1140 SPOEYVNDVRPQPPSPREGPLPAARPAAGATLIERAKTILSPGKXGVKDVAFGAGVEN 1199
Db 1134 SPOEYVNDVRPQPPSPREGPLPAARPAAGATLIERAKTILSPGKXGVKDVAFGAGVEN 1193
Qy 1200 PEYLTPOGGAAPHPHPPAPSPAFDNLYYMDODDPERGAPBSTFKGTPTAENPEYLGIDV 1259
Db 1194 PEYLTPOGGAAPHPHPPAPSPAFDNLYYMDODDPERGAPBSTFKGTPTAENPEYLGIDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 2
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiaxolylformamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'v', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F;658-680/Domain: transmembrane #status predicted <TMN>
F;723-988/Domain: protein kinase homology <KIN>
F;731-739/Region: protein kinase ATP-binding motif
F;71,191,263,535,576,634/Binding site: carboxyrate (Aen) (covalent) #status predicted
F;691/Binding site: phosphate (Thr) (covalent) #status predicted
F;758/Active site: Lys #status predicted
F;882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 86.5%; Score 5925; DB 1; Length 1260;
Best Local Similarity 86.6%; Pred. No. 1.1e-234;
Matches 1095; Conservative 56; Mismatches 103; Indels 10; Gaps 4;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTDMLKRLPASPETHDMLRHLVGGCQVQGNL 60
Db 4 MELAAWCRWGLLLALLPPGIAGTCTCTDMLKRLPASPETHDMLRHLVGGCQVQGNL 63

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 64 ELTYVPANASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNR 123

Qy 121 DPLNNTTPVT-GASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKQNL 179
Db 124 DPQNVAASPTGRTPEGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKQNL 183

Qy 180 ALTLIDNTRSRACHPCSPMKGSGESSDCOSLRTVTCAGGCARCKPLPTDCCHEQ 239
Db 184 APVDIDNTRSRACHPCCAPACKDNHGESPEDCQLTGICTSCACRCKRLPTDCCHEQ 243

Qy 240 CAAGCTGPKHSDDLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDDLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTAC 303

Qy 300 PYNVLSIDVGSCTLVCPHMQEVTAEQTCRCKSPCARVCVGLGMFNFTVSWFLRV 359
Db 304 PYNVLSIDVGSCTLVCPHMQEVTAEQTCRCKSPCARVCVGLGMFNFTVSWFLRV 359

Qy 360 PK-VSASHLEEFACCKIFSLAFPLPSFDGDPASNTAPLOPEOLQVFTLEETGYLYI 418
Db 357 ARAITSDNVEFDCCKIFSLAFPLPSFDGDPASNTAPLOPEOLQVFTLEETGYLYI 416

Qy 419 SAMPDLSLVSFONQVIRGLIHNGAYSLLTQGLISWGLRSLRSLGSLALIHNT 478
Db 417 SAMPDLSLVSFONQVIRGLIHNGAYSLLTQGLISWGLRSLRSLGSLALIHNA 476

Qy 479 HLCFVHTVPMDLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGCHGPGTQCVNC 537
Db 477 HLCFVHTVPMDLFRNPHQALLHTANRPEDE-CVGEGLACHQLCARGCHGPGTQCVNC 536

Qy 538 SFLRGQECVEECRVLQGLPREYVNAHRLCPCHPEQOPNGSVTCFGEADQCVACAHYK 597
Db 537 SHFLRGQECVEECRVLQGLPREYVNAHRLCPCHPEQOPNGSVTCFGEADQCVACAHYK 596

Qy 598 DPFCVACRCSGVKPDLSYMPIWKFPDEGACQPCINCTHSCVDLDKGCAPRQASPL 657
Db 597 DSSSVCARCSGVKPDLSYMPIWKFPDEGACQPCINCTHSCVDLDKGCAPRQASPL 656

Qy 658 TSIVSAVVGILLVVVLGWFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQ 717
Db 657 TFIATVEGVLILLVVVLGWFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQ 716

Qy 718 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEA 777
Db 717 MRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEA 776

Qy 778 YVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLLNWCQIAK 837
Db 777 YVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLLNWCQIAK 836

Qy 838 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 897

Db 837 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 896
Qy 898 ESIILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 957
Db 897 ESIILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 956
Qy 958 YMIWVKCMMIDSECRPRFRELVSFMRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLED 1017
Db 957 YMIWVKCMMIDSECRPRFRELVSFMRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLED 1016
Qy 1018 DDMGDLVDABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPESEBAP 1077
Db 1017 DDMGDLVDABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLPESEBAP 1076
Qy 1078 RSLPAPSEGAGSDVFDGDLGMAAKGLQSLPQRYSEDPTVLPSETDGYVAPL 1137
Db 1077 RSLPAPSEGAGSDVFDGDLGMAAKGLQSLPQRYSEDPTVLPSETDGYVAPL 1136
Qy 1138 TCSPOPEYVNPQVVRPQPPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAGGAV 1197
Db 1137 ACSPOPEYVNPQVVRPQPPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAGGAV 1196
Qy 1198 ENPEYLTPOGGAAPQPPHPPAFSPAFDNLYTWDQDPPERGAAPPSTFKTPTAENPEYI 1257
Db 1197 ENPEYLTPOGGAAPQPPHPPAFSPAFDNLYTWDQDPPERGAAPPSTFKTPTAENPEYI 1256
Qy 1258 DVPV 1261
Db 1257 DVPV 1260

RESULT 3
148161
P-185 precursor - golden hamster
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C;Accession: 148161
R;Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A;Reference number: 148161; MUID:94193007; PMID:7908275
A;Accession: 148161
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1254 <RES>
A;Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.3%; Score 5916.5; DB 2; Length 1254;
Best Local Similarity 86.3%; Pred. No. 2.4e-234;
Matches 1089; Conservative 66; Mismatches 98; Indels 9; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASCTCTDMLKRLPASPETHDMLRHLVGGCQVQGNL 60
Db 1 MELAAWCRWGLLLALLPPGIAGTCTCTDMLKRLPASPETHDMLRHLVGGCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVGYVLIHANNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNR 120

Qy 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKQNL 180
Db 121 DPLNNTTPVTGASPGGLRELQSLTEILKGGVLIQNPOLCYQDILWKDIFHKQNL 180

Qy 181 LTLIDNTRSRACHPCSPMKGSGESSDCOSLRTVTCAGGCARCKPLPTDCCHEQ 240
Db 181 LTLIDNTRSRACHPCSPMKGSGESSDCOSLRTVTCAGGCARCKPLPTDCCHEQ 240

R; Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A; Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A; Reference number: A60143; MUID: 85182650; PMID: 2985580
 A; Accession: A60143
 A; Molecule type: protein
 A; Residues: 740-744, 'X', 746-747 <RUS>
 R; Mroczkowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A; Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A; Reference number: A38023; MUID: 84191554; PMID: 6325948
 A; Contents: annotation; receptor activity
 A; Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R; Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A; Title: Functional independence of the epidermal growth factor receptor from a domain 2
 A; Reference number: A33331; MUID: 90003233; PMID: 2790960
 A; Contents: annotation; internalization signal
 C; Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C; Genetics:
 A; Gene: GDB: EGF
 A; Cross-references: GDB: 120610; OMIM: 131550
 A; Map position: 7p12.3-7p12.1
 C; Superfamily: epidermal growth factor receptor; protein kinase homology
 C; Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine
 F; 1-24/Domain: signal sequence #status predicted <SIG>
 F; 25-645/Domain: extracellular #status predicted <EXT>
 F; 75-300/Domain: EGF receptor extracellular domain repeat <BE1>
 F; 390-600/Domain: EGF receptor extracellular domain repeat <BE2>
 F; 646-668/Domain: transmembrane #status predicted <TM>
 F; 669-1210/Domain: intracellular #status predicted <INT>
 F; 710-975/Domain: protein kinase homology <KIN>
 F; 718-726/Region: protein kinase ATP-binding motif
 F; 999-1046/Region: coated-pit mediated internalization signal
 F; 1047-1210/Region: inhibitory
 F; 128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F; 745/Active site: Lys #status experimental

Query Match 46.0%; Score 3150; DB 1; Length 1210;
 Best Local Similarity 49.6%; Pred. No. 1.7e-121;
 Matches 631; Conservative 178; Mismatches 350; Indels 112; Gaps 23;

Qy	11	LLALLPPGAA--STOVCTGDMKRLPASPETHLMLRLHYQCVQVQGNLEITYLPTN	68
Db	14	LLAALCPASALEBKVCQGSNKLTLQGTGFEDHFLSLQRMFNCEVVLGNLEITYYQNR	73
Qy	69	ASLSFLQDIOEVQGYVLIHNNQVPLQRLIRIVRGTLQFEDNVALAVLQNGDPLNNTTP	128
Db	74	YDLSFLKTIQEVAGYVLIATLVERIPLENLIQIRGNMYVENSVALAVLSYD-----	126
Qy	129	VTGASPGQLRELQRLSRLTEILKGVLIQNPQLCYQDTILWKOIFHKNNQLALTLIDTNR	188
Db	127	---ANKTGLKELPMRNLIQELIHGAVRFSSNPALCNVESIQWRDITVSSDFLSNMSDFQNH	183
Qy	189	SRACHPCSPKCGSRCKGSESSDCQSLTRTVACGGCA-RCKGPLPTCCHEQCAAGCTGP	247
Db	184	LGSCQKCDPCSPNGSCMGAEENCQKLTIIQAQCCSGRCGRKSPSCCNCQCAAGCTGP	243
Qy	248	KHSDCLACHFNHSGICELCHPALVTYNTDTTFESMPNPEGRYTFGASCVTACPNYLSTD	307
Db	244	RESDCVCRKFRDEATCKDCTPLMLNPTTYQMDVNPCKYSGFATCVKCKPRNYYVTD	303
Qy	308	VGSCITVCLPHNOEVAEDGTQRCCKSKFCARVCYGLGNFNFTVFWLRVPKVSASHL	367
Db	304	HGSCVRACGADSYEM-EEDGVKCKCKCEGCRKVCNGIGI-GBFKDSL-----SINATNI	356
Qy	368	EEFAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVFTLEITCYLYISAWPDSLDP	427
Db	357	KHFNCTSSIGDHLIPLVAFRGDSFHTPPDDQELDKVKEITGFLLIQAWPENRTD	416
Qy	428	LSVFQNLQVIRILHNGAYSLLTQGLIGISWLGSLRLSRLGSLALHNNHLCFFVHTVP	487

RESULT 5

A53183

epidermal growth factor receptor precursor - mouse

C; Species: Mus musculus (house mouse)

C; Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999

C; Accession: A53183; A43818; S24942; A28941; S45325; I49643

R; Luettke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; Genes Dev. 8, 399-413, 1994

A; Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor

A; Reference number: A53183; MUID: 94170986; PMID: 8125255

A; Accession: A53183

A; Molecule type: mRNA

A; Residues: 1-1210 <LUE>

A; Cross-references: GB: U03425


```

RESULT 6
TUCHLV
epidermal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:Species: Gallus gallus (chicken)
C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720; A00643
R:Laux, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.2%; Score 3099.5; DB 1; Length 1223;
Best Local Similarity 48.3%; Pred. No. 1.9e-119;
Matches 630; Conservative 177; Mismatches 345; Indels 151; Gaps 26;
QY 8 RMGLLALLPPCAA-----STQVCTGTDMLRLPASPETHLDMLRHLVQGVQVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEERKVCQGNKTKUTQLGHVEDHFTSLQRMYNNECVLSNLE 72
QY 62 LTYLPTNASLSFLQIQEVGYLVAHNOVQVQLQRLIRVGTQLFEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTELTQIEVAGVYVLIANMVDVPLENQLIRGNVLYDNSFALAVLSNVH 132
QY 122 PLANVTPTVGSPGIGRELRLSLTEILKGGVLIQNPOLCVQDTILMKDIIPHKNQAL 181
DB 133 -MNKTQ-----GLRELPMKLEILNGGVKISNNPKLNMVDVLMNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGGCA-RCKGPLPTCCHEQ 239
DB 183 TVLDFASNLSSCPKCHPCTEDHCHGAGEQNCQTITKVIACQCSGRCKGRVPSCCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPGRESCLACRFRDQATCKDTCPLVLYNPITYQMDVNPPEGKYSFGATCVREC 302
QY 300 PNYLSTDVGSCVLPVLPNQVTAEDGTQRECKSKPCARVCYGLGMFNPNFTVSWFLRV 359
DB 303 PHNVVYVTDHGSCVRSNTDTTEV-EENGVRKCKKCDGLCKVCNGIG-----IGELKGI 355
QY 360 PKVSASHLEEFAGCKIFGSLAFLPESFDGDPASNTAPLQEPQLQVFTLEBITGYLIYS 419

```

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epiderma.

A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLOS>
A;Cross-references: GB:U07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126942)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Query Match 43.5%; Score 2981.5; DB 2; Length 1308;
Best Local Similarity 45.2%; Pred. No. 1.3e-114;
Matches 611; Conservative 187; Mismatches 375; Indels 179; Gaps 29;

QY 9 WGLLLALLPGAA-----STOVCTGTDMLKRLPASPTHLDMLRLHLYGCGVVOGNLELT 64
DB 8 WVVVSLVAAAGTVQPSDSQSVCAGTENKLSLSLEQQYRALRYKRYENCEVVMGNLEITS 67
QY 65 LPTNASISFLQDIOEVGYVLIHNOVQVPLQRLIRVGTOLPEQNYALAVLDNGDPLN 124
DB 68 IEHRDLSFURSREVGYVVALNOFRYLPLENRLIRGTUKLYEDRYALAI FLYNRKDG 127
QY 125 NTPPTVTCASPGGLRELQLRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQALTLI 184
DB 128 NF-----GLQELGKNLTEILNGVYVDQNKFLCYADTIHWQDIVRNPPSNLTIV 178
QY 185 DTVRSRACHPCSPMKGSRGWSSEDCQSLTRTVCAAGC-ARCKGPLPTDCHEQCAAG 243
DB 179 STNGSSGCRCHKSCCTG-RCWGPTEHNHCOTLRTVCAEQDCGRGYPVSDCHRECAAG 237
QY 244 CTGPKHSDCLACIHNHSGICELHCPALVTYNTDTESMPNPEGRYTFGASCVTACPNY 303
DB 238 CSGPKOTDCFACWNFDSGACVTCQPTQTFYNTPTTFOLEHNFNAKYTYGAFCKKCPHF 297
QY 304 LSTDVGSCTLVCPHLNHQVTAEDGTQRCERKSPCARVCYGLGMFNNFTVSFLRVPKVS 363
DB 298 V-VDSSSVCRACFPSSKMEV-EENGKMKPCCTDICKKACDGI-----TGLMSAQTV 349
QY 364 ASHLEFAGCKKIFGLSLAFIPESFDGDPASNTAPLQPEQLQVETLEITGYLYISAMP 423
DB 350 SSNIDRFNCTKINGNLIFLVTGTHGDPYNAIFAIDPEKLVNFRVREITGFLNIQSWPP 409
QY 424 SLPLSVFONQLOVIRGRLHNGAYSLTLOGLGTSLGLRLSLRELGLSLGLIHNTHLCFV 483
DB 410 NMTDFSVFSLNVTIGRVLVSGLSLLIKQGGITSLQFSLKISAGNIYITDNSNLCY 469
QY 484 HTVPWDLFRPNHQAHLHTANRPEDECVGEGLAACHQLCARHGCGPQTCVNCSPQLRG 543
DB 470 HTINWTLFTSTINQIRVIRDNRAENCTAEGVMCNHLCSDDGCGWGPQDCLSCRRFSR 529
QY 544 QECVECRVLQGLPREVYNARHCLPCHPECPQ-ONGSVTCFGEADOCVACAHYKPPFC 602
DB 530 RICIESCNLYDGFREFENGISICEVDCPOCEKMEDEGLLTCHGFGPDNCTKCSHFQDGP 589
QY 603 VAPCPSGVKPDLSYMPITWKFPEDEGACOPCPINCTHSCVDLDDKGC-----PAE 651
DB 590 VEKCPDGLQANF--IFKADPDRECHPCPNCTQCGNPTSHDCIYYPWTGHSTLPQH 647
QY 652 QRASPLTSIVSAVV-GILLYVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSG 710
DB 648 AR-TPL--IAAGVIGLFLIVVGLTFVAVVRRKSIK-KKRALRRFL-ETELVEPLTPSG 702
QY 711 AMPNQAMRLKTELAKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSKAN 770
DB 703 TAPNQALRLKTELKRVKVLGSGAGTGYKGIWVPEGETVPIPAIKILNETTPKAN 762
QY 771 KEILDAYVMAGVSPVSRLLGLTSTVOLVTLMPYCLLDHVRNRRGLRGSQDLN 830
DB 763 VEPMDALINASHDHPHLVRLVLLGVLCSPTIQLVTLMPHGLLLEYVHEHKDNGTSQLLN 822
QY 831 WCMQIAKGMYSLEDDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKV 890

DB 823 MCVOIAKGMYSLEERLVRDLAARNVLKSPNHVKITDFGLARLLSEGEKEYNADCGKM 882
QY 891 PIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDILLEKERLPOP 950
DB 883 PIKMALECIHYKFTHQSDVMSYGVTVWELMTFGGKPYDGIPTREIPDILLEKERLPOP 942
QY 951 PICTIDVYMWKCMWIDSECRPRFRFELVSEFSRMDARDPQVFFVIONED-LGPASPLDST 1009
DB 943 PICTIDVYMWKCMWIDSDSRKFKELAAAFSRMDARDPQVLYVIOGDDRMKLPSPNDK 1002
QY 1010 FYRSLLEDDDMGDLVDAEYLVPOQGFCCPDPAFGAGGMVHRRHSSTSRSGGDLTLGL 1069
DB 1003 FFQNLLEDEEDMDMAEEYLVLP-QAFNIPPP-----IYTSRARIIDNRS-----EIGH 1050
QY 1070 EPSEEEAPRS-----PLAP-SEGAGSDVDFDGLMGMA 1100
DB 1051 SPPPAYTPMSGNOFVYRDGGFAAEGQSVSPYRAPRTSTIPEAPVAQAGATAEIFDDSCCNGT 1110
QY 1101 AKGLQSLPTHDPSPLOQYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRP 1153
DB 1111 LRKPAPHVOEDSSTQRYSDAPTVFAPERSPRGELDEEGYTMPRDKPKOEYLYNPVE--- 1167
QY 1154 QPPSPREGPLPAARPAAGATLERAKTILSPGXNGVVKDFAFGGAIVENPEYLTPOGCAAPQ 1213
DB 1168 -----ENPVSRR-----KNGDLQ-----ALDNEYHNASNG----- 1194
QY 1214 HPPPA-----FSPAFDNLYYWDQDPPPERGA- 1238
DB 1195 -PKAEDEVYNEPLYLNTFANTLGKAEYLNKNTILSMPEKAKAFDNDPYWNHSLPPRSTL 1253
QY 1239 -PPSTFKGTPT-----AENPEYL 1255
DB 1254 QHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related transmembrane
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Zaulf, F.; Telling, A.; Rober
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu 1
Oncogene 6, 73-80, 1991
A;Reference number: S06142; MUID:90015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophorus
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
C;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; ty
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 39.0%; Score 2672; DB 1; Length 1166;
Best Local Similarity 44.9%; Pred. No. 5e-102;

F:229/Active site: Lys #status predicted

	Query Match	25.8%	Score 1766.5;	DB 1;	Length 698;
	Best Local Similarity	52.2%	Pred. No. 2.7e-65;		
	Matches 374;	Conservative 80;	Mismatches 137;	Indels 125;	Gaps 18
Qy	584	GPEADQCACAHYKDPFCVARCSPGVKPDLSYMP	KFPDDEEGACQPCTINCHSCVDL	643	
Db	60	GP--DHCMKAHFIDGPHCVKACPAGVLGENDTL-	VWKYADANAVCOLCHPNCTRGRCKGP	116	
Qy	644	DDKGCPAEQRASPLTSTISVAW-V	GILLVVVGUVFGILIKRQQKIRKYTMRRLLQETEL	702	
Db	117	GLEGPC---NGSKTPSIAAGVVGGLCLVVGVLGI	LRLRR-HIVRKRTLRLLLQEREL	172	
Qy	703	VEPLTPSGAMPNOAQMRIKETELRKVKVLGSAGF	TVYKGIWIPDGENVKIPIVAIKULR	762	
Db	173	VEPLTPSGEAPNQHLRIKETEFKKVKVLGSAGF	TVYKGLWIPEGEKVKIPIVAIKELR	232	
Qy	763	ENTSPKANKEILD EAYVMAGVGS PVSRLLGICLT	STVOLTMPYPYCGLLDHVENRCR	822	
Db	233	EATSPKANKEILD EAYVMAS VDNPHVCRLLGICLT	STVOLITQLMPPYCGLLDYIREHKDN	292	
Qy	823	LGSQDLLNWCMIQAKGMSYLEDVRLVRDLAARNV	LKSNHNKVTDFGLARLLDIDETE	882	
Db	293	IGSQYLLNWCVOIAKNMYLEERLVHRDLAARNV	LKTPHVKITDFGLAKLLGADEKE	352	
Qy	883	YHADGGKVPIKMMALESILRRFRTHQSDVMSYG	VTVMELMTFGAKPYDGIPAREIPDLLE	942	
Db	353	YHAEGGKVPIKMMALESILHRIYTHQSDVMSYG	VTVMELMTFGSKPYDGIPASEISSVLE	412	
Qy	943	KGERLPQPTICTIDVYMIMVKCWMIDSECRPR	FRELVSFGRMARDPQRFVVIQ-NEDLG	1000	
Db	413	KGERLPQPTICTIDVYMIMVKCWMIDASRPKRE	LIAEFASKWARDPPRYLVIQGDERMH	472	
Qy	1002	PASPLDSTFYRSLLDDMDGLVDAAEVLPQGFCD	PAPGAGGMVHHRHSSTSRG	1060	
Db	473	LPSSTD SKFYRTLMEEEDMEDIVDAEYLVPHQGF	-----NSPST---	513	
Qy	1062	GGDLTLGLSEEBEAPRSPL-----APSEGA	GSVDVFDGDLGWGAAGLQSLPTHDPSPLO	1110	
Db	514	-----STRPLSSLSATSNNSATNCID-----	RNCGHPVREDTSFVQ	550	
Qy	1117	RYSEDPTVLPSET--DGYVAPLTCSPQEYVNPQ	DRVPPSPREGPLPAARPAGATILE	1170	
Db	551	RYSSDPTGNFLSESIDGFL-----PAPEYVNQ	--LMPKXPS-----	585	
Qy	1175	RAKTLSPGKNGVKWDVF-----AFGGAVEN	PEYLTPOGGAAPOPHPPAFP	1210	
Db	586	----TAMVQNQIYNNISILTAISKLPMDSRVQNSH	STAVDNFYL-----NTNQSPLA	633	
Qy	1220	SPAFDNLIIYNDQ-----DPPE-----RGAPP	STFKTPTAENDEYLGLDVP	1260	
Db	634	KTFPESSPYIQSGNHQINDLNPDYQODFLNP	ETKPNGLLKVPAAENDEYLRVAAP	689	

RESULT 12

TWUHH
 Protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
 C)Species: avian erythroblastosis virus
 C/Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
 C/Accession: A00644; A38022
 R/Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
 Cell 35, 71-78, 1983
 A/Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
 A/Accession number: A00644; MUID:8402653; PMID:6313229
 A/Reference: A00644
 A/Molecule type: DNA
 A/Residues: 1-604 <YAM>
 A/Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
 R/Debuire, B.; Henry, C.; Benalassa, M.; Biserte, G.; Claverie, J.M.; Saule, S.;
 Science 224, 1456-1459, 1984
 A/Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new
 A/Reference number: A38022; MUID:84223957; PMID:6328658

A:Accession: A38022
A:Molecule type: DNA
A:Residues: 1-28,'W',30-139,'F',141-145,'V',147-152 <DB>
A:Cross-references: GB:K02006
C:Genetics:

C:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:130-395/Domain: protein kinase homology <KIN>
F:138-146/Region: protein kinase ATP-binding motif
F:165/Active site: Lys #status predicted

Query Match 24.9%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 9e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

QY 593 CAHYKDPFCVACRGKVPDLVWVGVFGILIKRRQKIRKYTMRLRLQETELVPTSGA 652
DB 3 CAHFDGPHCVKACVAGVLGENDTL-VKRYADANAVCQLCHPNCNTRGCKGPGLEGCP--- 58
QY 653 RASPLTSIVSAVV-GILLVWVGVFGILIKRRQKIRKYTMRLRLQETELVPTSGA 711
DB 59 NCSKTPSTAAGVVGGLLVVGLGILYLRRL-HIVKRTLRLLQERELVEPLTPSGE 117
QY 712 MPNQAMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANK 771
DB 118 APQAHLRIKETEFKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANK 177
QY 772 EILDAYVMAGVGSVVSRLGICLTSTVQLTQMPYGCGLLDHVRNRRGRSGOGLNW 831
DB 178 EILDAYVMASVDNPHVCRLLGICLTSTVQLTQMPYGCGLLDYIREHKDNGSQVLLNW 237
QY 832 CMQIAGMSYLEDVRLVHRDLAARNVVKVSPNHVKITDFGLARLLDIDETEHADGKVP 891
DB 238 CVQIAGMNYLEERLVRDLAARNVVKVTPQHVKITDFGLAKLLGCADEKEYHAEGKVP 297
QY 892 IKWMALESILRRRFTQSDVMVSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPP 951
DB 298 IKWMALESILHRIYTHQSDVMVSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
QY 952 ICTIDVYIMVVKWIMDSECRPRFRELVSFEMARDPQFVVIO-NEDLGASPLDSTF 1010
DB 358 ICTIDVYIMVVKWIMDASRPKRELIAEFKQARDPPRYLVIOGDERMHLPSPTDSKF 417
QY 1011 YRSLLEDDMDGLVDAEYLVVQQGFPCPDPAFGAGVHHRHRSSTRSGGDLTLGLE 1070
DB 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1071 PSEERAPRPL-----APSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSLPQRYSEDPTVP 1125
DB 450 -----SRTPLLSLSLSSATNSNATNCID-----RNCQGHFVREDSFVQRYSDPTGN 495
QY 1126 LPSET--DGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTLPCK 1183
DB 496 FLEESIDDDGL-----PAPEVYNO--LMPKPKSTAM----- 524
QY 1184 NGVVKDVFAP-----GGAVENPEYLTPOGGAAPQHPPPAPFAPD 1224
DB 525 --VQNIYNFISLTAKSLPMDSRYQNSHSTAVDNPEYL-----NTNQSPKAKTVEF 574
QY 1225 NLYVMDQDPPPERGAPPSTFKGTPTAENPEY 1254
DB 575 SSPYMTQSGNHQ-----INLDNPDY 594

RESULT 13

S35745
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
A:Accession: S35745
R:Venustroem, B.
submitted to the EMBL Data Library, March 1993

A:Reference number: S35743

A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match 24.0%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.6e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 584 GPEADQCVACAHYKDPFCVACRGKVPDLVWVGVFGILIKRRQKIRKYTMRLRLQETEL 643
DB 1 GP--DHCMKCAHFDGPHCVKACVAGVLGENDTL-VKRYADANAVCQLCHPNCNTRGCKGP 57
QY 644 DDGCPAQARASPLTSIVSAVV-GILLVWVGVFGILIKRRQKIRKYTMRLRLQETEL 702
DB 58 GLEGCP---NGSKTPSTAAGVVGGLLVVGLGILYLRRL-HIVKRTLRLLQEREL 113
QY 703 VEPLTPSGAMPNQAMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLR 762
DB 114 VEPLTPSGEAPNQAHRLRIKETEFKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLR 173
QY 763 ENTSPKANKEILDEAYVMAGVGSVVSRLGICLTSTVQLTQMPYGCGLLDHVRNRRGR 822
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLTQMPYGCGLLDYIREHKON 233
QY 823 LGSQDLNWCQIAGMSYLEDVRLVHRDLAARNVVKVSPNHVKITDFGLARLLDIDETE 882
DB 234 LGSQDLNWCQIAGMNYLEERLVRDLAARNVVKVTPQHVKITDFGLAKQLGCADEKE 293
QY 883 YHADGKVPKIKWMALESILRRRFTQSDVMVSYGVTVWELMTFGAKPDGIPAREIPDLLE 942
DB 294 YHAEKGKVPKIKWMALESILHRIYTHQSDVMVSYGVTVWELMTFGSKPYDGPASEISSVLE 353
QY 943 KGERLPQPPICTIDVYIMVVKWIMDSECRPRFRELVSFEMARDPQFVVIO-NEDLG 1001
DB 354 KGERLPQPPICTIDVYIMVVKWIMDASRPKRELIAEFKQARDPPRYLVIOGDERMH 413
QY 1002 PASPLDSTFYRSLLEDDMDGLVDAEYLVVQQGFPCPDPAFGAGVHHRHRSSTRSG 1061
DB 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 454
QY 1062 GGDLTLGLEPSEERAPRPL-----APSEGAGSDVFDGLGMGAAGKGLQSLPHTDPSLPQ 1116
DB 455 -----SRTPLLSLSLSSATNSNATNCIDRNGG-----H----- 481
QY 1117 RYSEDPVTPLPSETGYVAPLTCSPQPEYVQPDVVRPQPPSPREGPLPAARPAGAT-LER 1175
DB 482 -----PVREDGFL-----PAPEVYNO--LMPKPKSTAMVQNIYVYISLTAKS 523
QY 1176 AKTLSPGKNGVVKDVFAPGGAIVENPEYL 1203
DB 524 LPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14

GOFPFE

epidermal growth factor receptor - fruit fly (Drosophila melanogaster)

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Drosophila melanogaster

C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999

C:Accession: A00640; A38021

R:Livneh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.

Cell 40, 599-607, 1985

A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding

A:Reference number: A00640; MUID:85124611; PMID:2982499

A:Accession: A00640
A:Molecule type: DNA
A:Residues: 1-1330 <LIV>
A:Cross-references: EMBL:K03054
R:Wadsworth, S.C.; Vincent III, W.S.; Bilodeau-Wentworth, D.
Nature 314, 178-180, 1985
A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A:Reference number: A38021; MUID:85137938; PMID:2983232
A:Accession: A38021
A:Molecule type: DNA
A:Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <WAD>
A:Cross-references: EMBL:X02293; NID:g7922; PIDN:CA26157.1; PID:g929565
C:Comment: This sequence is tentative because the introns have not been identified.
C:Genetics:
A:Gene: FlyBase:Egfr
A:Cross-references: FlyBase:FBgn0003731
A:Map position: 2 57F
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP, autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F:1-732/Domain: extracellular #status predicted <EX>
F:733-764/Domain: transmembrane #status predicted <TM>
F:765-1330/Domain: intracellular #status predicted <INT>
F:808-1072/Domain: protein kinase homology <KIN>
F:816-824/Region: protein kinase ATP-binding motif
F:122-300, 324, 363, 518, 688, 695, 700/Binding site: carbohydrate (Asn) (covalent) #status pr
F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:843/Active site: Lys #status predicted
F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predict

	Query Match	24.0%	Score 1643.5	DB 1	Length 1330
	Best Local Similarity	29.7%	Pred. No. 5.2e-60		
	Matches 412	Conservative 181	Mismatches 415	Indels 377	Gaps 39
Qy	80	VGYVLIANNQVRQVQLRLIRIVRGTOLF-----EDNYALAVLNDGDPFNNTTPTVTGASP	134		
Db	38	ITNYIVIGLDLPCTLSYRLQIIRGRTLFLSLVEEEKYALFV-----TY	81		
Qy	135	GGRLQLRLSLTEILAGGVLIQRNPOLCYODTILWKDIFHKNNQLALTLIDNRSRACHP	194		
Db	82	SKMYTLEIPDLRDVLNGQVGFHNHYNLCHMRTIQMSEIVSNGTDAYNYDFTAPERCPK	141		
Qy	195	CSPMCKGSRWCWGESSDCQSLRTVCAGGCA--RCKGPLPTDDCCEQCAAGCTGPKHSDC	252		
Db	142	CHESCTHG-CWGEPPKNCQKFSKLTCSPOCAGGRCYGPKECCCHLCFAGCGCTGTQDC	200		
Qy	253	LACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACPNYLSLTDVGSC	312		
Db	201	IACKNFFDEAVSKEECPMPKRYNPTTYVLETPNPEGYATGATCVKECP-CHLLRDNGACV	259		
Qy	313	LVCPLHNQEVTAEDGTQRCEKSKPCARVCYGLGMFNFTVSFWLRPKVSASHLEEFAG	372		
Db	260	RSCPODMKGG-----CVPNGPCPKTCF-----VTVLHAGNIDSPN	300		
Qy	373	CKKIFGSLAFLPESFDG-----DPASNTA-----PLQEQLOQVFETLEEITGYLYTSAMPDS	424		
Db	301	CTVIDGNIRILQDTFSGFDQVYANYTMGPRIPLDPERREVFTSVKEITGYLNTGTHPQ	360		
Qy	425	LPDLSVFQNLQVIRGLHNGAY-SLTQGLGTSWLGLRSLRELGSGLALIHNNTHLCFV	483		
Db	361	FRNLSYFRNLEITHGRLQMESMPAALAVKSSLSYSLEMRNLKQISSGSSVVIQHNRDLGV	420		
Qy	484	HTVPMDOLFNRPHQALLHTANRPEDEC-----	510		
Db	421	SNIRPATOKEPEQKVWVNNENLADLCGKFLTILISVQHNIIMHIFAICREKNHLLGSV	480		
Qy	511	-----	510		
Db	481	QRGRLLGSHGSPVYQLQELQFOWHLHRLRLWLIQVINSITQDKSNEHQLTDACYSPSVPT	540		
Qy	511	-----	512		
Db	541	SLTIERARYAIQSAGLAMELEQITARSASNRHRSKTLPAEGRQVPRWVFLGVCASARAGIA	600		

C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.9%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

Qy 584 GPEADOCVACAHYKOPFCVACPCSPKPDLSYMPIWKPFDEEGACQPCPINCTHSCVDL 643
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGP 57
Qy 644 DDKGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 702
Db 58 GLEGCP---NGSKTPSIAAGVWGGLLCLVVVGLIGLRLRR-HIVKRTLRLLQEREL 113
Qy 703 VEPLTPSGAMPNQAOQRILKETELRKVKVLGSGAFGVYKGIWIDGENVKIPVAIKVL 762
Db 114 VEPLTPSGEAPNQAHRLILKETEFKVKVLGFGAFGVYKGLWPEGEKVITIPVAIKEL 173
Qy 763 ENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRR 822
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQIMPYGCLLDYIREHKON 233
Qy 823 LGSQDLLNMCQIAKGMVLEVDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 882
Db 234 IGSQYLLNMCVQIAKGMNILEERHLVHRDLAARNVLKTPQDVKITDFGLAKQLGADEKE 293
Qy 883 YHADGKVPKIMWALESIILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLE 942
Db 294 YHAEGGKVPKIMWALESIILHRIYTHOSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353
Qy 943 KGERLPOPPICITIDVYIMVYKMWIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLG 1001
Db 354 KGERLPOPPICITIDVYIMVYKMWSDADSRPKFRELIAEFKWARDPPRYLVIQGDERMH 413
Qy 1002 PASPLDSTFYRSLLDDMDGLVDAEYLVPOQGFPCDPAPAGAGMVHHRHRSSTRSG 1061
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1062 GGDLTLGLEPSEEAAPRSPL-----APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLO 1116
Db 455 -----SRTPLLSLSATSNSNATNCIDRNGG-----H----- 481
Qy 1117 RYSEDPTVPLPSETDGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER 1175
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKKESTAMVQNIYNIISLTAISK 523
Qy 1176 AKTLSPGKNGVVKDVFAPGGAVENPEYL 1203
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544

Search completed: July 22, 2003, 09:09:31
Job time : 31.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds
(without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	6730	98.8	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5963	87.6	1257	1 ERB2_RAT	P06494 rattus norv
3	5941.5	87.3	1254	1 ERB2_MESAU	Q05053 mesocricetu
4	3158	46.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3131	46.0	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	3003.5	44.1	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2984	43.8	1308	1 ERB4_RAT	Q62956 rattus norv
8	2715.5	39.9	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2430.5	35.7	1342	1 ERB3_HUMAN	P21880 homo sapien
10	2359.5	34.7	1339	1 ERB3_RAT	Q62799 rattus norv
11	1976	29.0	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1614	23.7	703	1 EGFR_CHICK	P13387 gallus gall
16	1311	19.3	1323	1 LT23_CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	733	10.8	1363	1 ILPR_BRALA	O02466 branchiosto
19	718	10.5	1382	1 INSR_HUMAN	P06213 homo sapien
20	713	10.5	1383	1 INSR_RAT	P15127 rattus norv
21	712.5	10.5	1372	1 INSR_MOUSE	P15208 mus musculu
22	699	10.3	1300	1 IRR_MOUSE	Q9wt14 mus musculu
23	694	10.2	1607	1 MIPR_LYNST	Q25410 lymnaea sta
24	690	10.1	1297	1 IRR_HUMAN	P14616 homo sapien
25	685.5	10.1	1297	1 IRR_CAVPO	P14617 cavita porce
26	684	10.0	1477	1 HTK7_HYDAT	Q25197 hydra atten
27	651	9.6	1367	1 IGLR_HUMAN	P08089 homo sapien
28	644	9.5	1390	1 INSR_ADAE	Q93105 aedes aegypt
29	641	9.4	1373	1 IGLR_MOUSE	P60751 mus musculu
30	637.5	9.4	1370	1 IGLR_RAT	P24062 rattus norv
31	617	9.1	2146	1 INSR_DROME	P09208 drosophila
32	605	8.9	987	1 EPB4_HUMAN	P54760 homo sapien
33	591.5	8.7	984	1 EPB1_CHICK	Q07494 gallus gall

ALIGNMENTS

RESULT 1

ERB2_HUMAN	589.5	8.7	977	1	EPB2_MOUSE	Q03145
34	589.5	8.7	977	1	EPB2_MOUSE	Q03145 mus musculu
35	588	8.6	1114	1	RET_HUMAN	P07949 homo sapien
36	587	8.6	987	1	EPB4_MOUSE	P54761 mus musculu
37	584.5	8.6	976	1	EPB2_HUMAN	P29317 homo sapien
38	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
39	579	8.5	902	1	ERBB_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu
ERB2_HUMAN	AC	ERB2_HUMAN	STANDARD;	PRT;	1255 AA.	
DT	13-AUG-1987	(Rel. 05, Created)				
DT	13-AUG-1987	(Rel. 05, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)					
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell					
DE	surface receptor HER2) (MLN 19).					
DN	ERBB2 OR HER2 OR NGF OR NEU.					
OS	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86118663; PubMed=3003577;					
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,					
RA	Saito T., Toyoshima K.;					
RT	"Similarity of protein encoded by the human c-erbB-2 gene to					
RT	epidermal growth factor receptor.";					
RL	Nature 319:230-234 (1986).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86070181; PubMed=2999974;					
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,					
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,					
RA	Francke U., Levinson A., Ullrich A.;					
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor					
RT	shares chromosomal location with neu oncogene.";					
RL	Science 230:1132-1139 (1985).					
RN	[3]					
RP	SEQUENCE OF 737-1031 FROM N.A.					
RX	MEDLINE=86016729; PubMed=2995967;					
RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;					
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the					
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a					
RT	human salivary gland adenocarcinoma.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).					
RN	[4]					
RP	VARIANTS VAL-654 AND VAL-655.					
RX	MEDLINE=93194196; PubMed=8095488;					
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;					
RT	"Characterization of a new allele of the human ERBB2 gene by allele-					
RT	specific competition hybridization.";					
RL	Genomics 15:426-429 (1993)					
CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,					
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A					
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-					
CC	ALPHA AND AMPHIREGULIN.					
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein					
CC	tyrosine phosphate.					
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS					
CC	(POTENTIAL).					
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.					

11 22 12:39:05 2003

PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

RESIDUES (BY SIMILARITY) FOURS ALLELES DUE TO THE VARIATIONS IN

POLYMORPHISM: THERE ARE FOURS ALLELES B1 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;

POSITIONS 654 AND 655 ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.012.

OF 0.782; ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.

ALLELE B4 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.

SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

IT IS PRODUCED THROUGH A COLLABORATION

BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION

AT DES HAIES, FRANCE. THERE ARE NO RESTRICTIONS ON ITS

USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS NOT FOR COMMERCIAL

PURPOSES. THIS STATEMENT IS NOT REMOVED. (See <http://www.isb-sib.ch/announce/>)

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.

OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH.


```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.6%; Score 5963; DB 1; Length 1257;
Best Local Similarity 87.4%; Pred. No. 9.3e-313;
Matches 1099; Conservative 51; Mismatches 105; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHDMLRHLRYGCGVQGNL 60
Db 1 MELAARCRWGLLLALLPPGIAGTQVCTGDMKLRLPASPETHDMLRHLRYGCGVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVQPLQRLIRVGTQQLFEDNYALVDNG 120
Db 61 ELTYVPANASLSFLQDIOEVQGYMLIAHNOVKRVPQLRLIRVGTQQLFEDKYALVDNR 120
Qy 121 DPLNNQYIKANSKF-IGITELQLRLSLTEILKGGVLIQBNPOLCVQDTILWKDIFHKNNQL 179
Db 121 DPQDNVAASTPGRTPGELRELQLRLSLTEILKGGVLIIRGNPOLCYQDMVLWKDVFRRKNNQL 180
Qy 180 ALTLIDNRSRACHPCSPMKGSRGWSESDCSLTRVTCAGGACRCKGLPTDCCHEQ 239
Db 181 APVDIDNRSRACHPCPACADKNHCWGESPEDCQLTGTCTSCARCKGLPTDCCHEQ 240
Qy 240 CAAGCTGPKSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTAC 299
Db 241 CAAGCTGPKSDCLACLHFNHSGICELHCPALVYNTDTFESMENPEGRTYFGASCVTTC 300
Qy 300 PYNLYSTDVGSCTLVCPHNOEVAEDGTQCEKSPCARVCYGLGMEHLRVRVATSA 359
Db 301 PYNLYSTEVGSCTLVCPNNOEVAEDGTQCEKSPCARVCYGLGMEHLRGARAITSD 360
Qy 360 NIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLQVETLEEITGYLIYISAWPDSL 419
Db 361 NVQEFDCCKKIFGSLAFLPESFDGDPSSGIAPLPEQLQVETLEEITGYLIYISAWPDSL 420
Qy 420 PDLVSFQNLQVIRIILHNGAYSILTLOGLGISWLGRLSLRELGSGLALIHHTHLCFVHT 479
Db 421 RDLVSFQNLRIIRGRIILHDGAYSILTLOGLGISHLSRLSLRELGSGLALIHNAHLCFVHT 480
Qy 480 VPMDLFRNPHOALLHTANREDE-CYEGELACHOLCARHCWGPGTQCNCSCQFLRGQ 538
Db 481 VPMDLFRNPHOALLHSGNRPEEDLVCSSGLVNSLCARHCWGPGTQCNCSCQFLRGQ 540
Qy 539 ECVEECRVLOGLPREYVNRARCLCHPECPQNGSVTCFGEADQCVACAHYKDPFCA 598
Db 541 ECVEECRVWGLPREYVSDRCLCHPECPQNSSETCFGEADQCAAHYKDKSSCVA 600
Qy 599 RCPGSGVPDLSYMPIWKPDEEGACQPCINCHTSCVDLDDKGCAPQASPLTSIVSAV 658
Db 601 RCPGSGVPDLSYMPIWKYPDEEGICQPCINCHTSCVDLDERGCPAQASPLTVIATV 660
Qy 659 VGILLVVVGVFGILIKRROOKIRKYMRLLOETELVEPLTSGAMPNQAOIRILKET 718
Db 661 VGULLFLLVVVGVGILIKRRQKIRKYMRLLOETELVEPLTSGAMPNQAOIRILKET 720
Qy 719 ELRKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAVG 778
Db 721 ELRKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYVMAVG 780
Qy 779..SPYVSRLLGLICTSTVQLTQMPYGCLLDHVREHGRGLSGODLLNMCVQIAKMSYLED 838
```

```
Db 781 SPYVSRLLGLICTSTVQLTQMPYGCLLDHVREHGRGLSGODLLNMCVQIAKMSYLED 840
Qy 839 VRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPYIKMALESILRRR 898
Db 841 VRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPYIKMALESILRRR 900
Qy 899 FTHOSDVWSGYVTWELMTFCGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 958
Db 901 FTHOSDVWSGYVTWELMTFCGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMWKC 960
Qy 959 WMIDSECRPRELVSFSEFMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDMGDLV 1018
Db 961 WMIDSECRPRELVSFSEFMRARDPQRFVVIQNEDELGPSSPMDSTFYRSLLEDDDMGDLV 1020
Qy 1019 DAESYLVPOQFFCFDPAPGAGGWHHRSSSTRSGGGDLTLGLEPSEEAAPRPLAPS 1078
Db 1021 DAESYLVPOQFFCFDPPTGPGTAHRRSSSTRSGGGELTLGLEPSEEGPRPLAPS 1080
Qy 1079 EGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVLPSPSETDGYVAPLTCSPOPE 1138
Db 1081 EGAGSDVFDGDLGMAAGVTKGQSLPSPLQRYSEDPVLPSPSETDGYVAPLTCSPOPE 1140
Qy 1139 YVNPQDVRPPSPREGPLPAARPAAGATLERAKTLPCKNGVVKDVFAGGAVENPEYLT 1198
Db 1141 YVNPQEQVQPLTPEGLPPVRPAGATLERPKTLPCKNGVVKDVFAGGAVENPEYLV 1200
Qy 1199 PQGAAAPQHPPPAFSPAFDNLVYWDQPPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255
Db 1201 PREGTASPPHPSPAFSPAFDNLVYWDQNSSEQPPSPNFECTPTAENPEYLGLDVVPV 1257

RESULT 3
ERB2_MESAU
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60533;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) - (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OK NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF. TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



```
Db 536 CNUYDGEFREFENGSCVCECDPOCEKMEGDGLLTCHGPGPNCTKCSHFHKGPNVCNVEKCPD 595
Qy 603 GVKPDLSSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGC-----PAEQRASPL 651
Db 596 GLOGANSF--IFKYADPDRECHCPNCTQCGNGPTSHDCIYYPWTCGHSTLPQHAR--TPL 652
Qy 652 TSIVSAVV--GILLVVVLGVVGLIIRKROQIKIRKYMRRLLQETELVEPLTPSGAMPNOA 710
Db 653 --IAAGVIGGLTFLVTVGLTFVAVVRKSIK-KKRALRREL-ETELVEPLTPSGTAPNOA 708
Qy 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRENTSPKANKETLDE 770
Db 709 QLRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRENTSPKANKETLDE 768
Qy 771 AYVMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCILLDHVRENRLGRLSQDLLLLNCWQIA 830
Db 769 ALIMASMDPHVRLVLLGVCLSPITQLVTLQMPHGLLEYVHEHKDNIGSQDLLLLNCWQIA 828
Qy 831 KGSYLEDVRLVHRDLAARLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMA 890
Db 829 KGMYLEERLRLVHRDLAARLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMA 888
Qy 891 LESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTID 950
Db 889 LECIHYRKTHQSDVMSYGVYVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTID 948
Qy 951 VYMIMVKCWMIDSECRPRELSEFSRMARDPQRFVVIQNE-LGPASPLDSTFYRSL 1009
Db 949 VYMIMVKCWMIDSECRPRELSEFSRMARDPQRFVVIQNE-LGPASPLDSTFYRSL 1008
Qy 1010 EDDMDGLDVAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEE 1069
Db 1009 DEEDLEMDMAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEE 1056
Qy 1070 APRS-----FLAP--SEGAGSDVFDGDLGMAAGLQ 1100
Db 1057 TMSGNQFYVRDGGFAAEQGVSVYRAPSTIPEAPVACQATAEIEDDSCCNTLKRPA 1116
Qy 1101 LPHDPSPIQRYSDDTVPPLPS-----ETDGYVAPLTCSPQPEYVQNPQDVRPQPPSPR 1153
Db 1117 PHVQEDSSQRYSDPTVPAPERSPRGELDEEGYMTMRDKPKQEVLPVPE----- 1167
Qy 1154 EGPLPAARAGATLERAKTLSPKNGVVKVDVAFGAVENPEYLTQGGNAOPHPHPPA- 1212
Db 1168 ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAE 1199
Qy 1213 -----FSPAFDNLVYWDQDPPPERGA--PPSTF 1237
Db 1200 DEYVNEPLVNTFANTLGAEYLNKLNLSMPEKAKAFDNPDIWNHSLPRSTLQHPDYL 1259
Qy 1238 KGTP-----AENPEYL 1249
Db 1260 QEYSTKYFYKQNGRIRPPIVAENPEYL 1285
RESULT 7
ERB4_RAT
ID ERB4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAD08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_Pkinase.
CC InterPro; IPR004019; YLP_motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP_2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```

FT	DOMAIN	186	334	CYS-RICH.	Qy	357	TSANIOEFAGCKKIFGSLAFLESFDPASNTAPLOEQVQVFEETEEITGLYLVISAMP	416
FT	DOMAIN	496	633	CYS-RICH.	Db	349	DSSNIDKFNCTKINGNLIFLVGTIHGDPYNAIDAIPEKLVNRTVREITGLFNLQTPW	408
FT	NP_BIND	724	732	PROTEIN KINASE	Qy	417	DSLPLSVFQNLQVIRGRIHLNGAYSILTQGLGISWLGRLSLRELGSGLAIHHNTHICF	476
FT	ACT_SITE	751	751	ATP (BY SIMILARITY).	Db	409	PNMTDFSVFNLVTIGRVLYSGLSLILKQOGITSLQFQSLKEISAGNIVITONSLEY	468
FT	DISULFID	843	843	BY SIMILARITY.	Qy	477	VHTVPDOLRNPHOALLHTANRPEDECVGGLACHOLCARGHGWGPGTQVNCVSQFLR	536
FT	DISULFID	189	197	BY SIMILARITY.	Db	469	YHTINWTTFLSTVNRIVIRNRAENCTAEGMYCNHLCSNDGCGPGPDQCLSCRRFSR	528
FT	DISULFID	193	205	BY SIMILARITY.	Qy	537	GQECVEECRVLQGLPREYVNAHCLPCHPECPQ-QNGSVTCFGEADOCVACAHYKDPFF	595
FT	DISULFID	213	221	BY SIMILARITY.	Db	529	GKICIESCNLYDGEFEFENGSI-CVECDSCQEKMEDGLLTCHGCPDNCTKSHFQDGN	588
FT	DISULFID	217	229	BY SIMILARITY.	Qy	596	CVARCPGVPDLSYMPIMKPPDEEGACQPCPINCTHSCVDLDDKGC-----PA	644
FT	DISULFID	230	238	BY SIMILARITY.	Db	589	CVEKCPDVLQGANF--IFKYADQDRCHPCHPNCCTQCGNPTSHDCIYYPMWGHSTLPQ	646
FT	DISULFID	234	246	BY SIMILARITY.	Qy	645	EORASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTHRRLLQETELVEPLTPS	703
FT	DISULFID	249	258	BY SIMILARITY.	Db	647	HAR--TPL--IAAGVIGGLFILVINALTFVAVVRRKSIK-KKALRRFL-ETELVEPLTPS	701
FT	DISULFID	262	289	BY SIMILARITY.	Qy	704	GAMPNOAQRILKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKA	763
FT	DISULFID	293	304	BY SIMILARITY.	Db	702	GTAPNOAQLRIKETELRKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKA	761
FT	DISULFID	308	323	BY SIMILARITY.	Qy	764	NKEILDYAVMAGVSPVSRLLGICLTSTVQLTQVLMVYVCLLDHVRNRLGSGDQLL	823
FT	DISULFID	326	330	BY SIMILARITY.	Db	762	NVEFMDALIMASVDHPLVRLVLLGVCLSPITQVLTQVLMVYVCLLDHVRNRLGSGDQLL	821
FT	MOD_RES	503	512	PHOSPHORYLATION (AUTO-)	Qy	824	NWCMOIAKGSYLEDVRLVHRDLAARNVVKVSPNHVKITDGLARLLDIDETEHYADGCK	883
FT	MOD_RES	507	520	PHOSPHORYLATION (AUTO-)	Db	822	NWCQIAKGMVLEERLVRDLAARNVVKVSPNHVKITDGLARLLDIDETEHYADGCK	881
FT	MOD_RES	523	532	PHOSPHORYLATION (AUTO-)	Qy	884	VPIKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQ	943
FT	MOD_RES	536	552	PHOSPHORYLATION (AUTO-)	Db	882	MPKWMALCICHYRKFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQ	941
FT	CARBOHYD	555	569	N-LINKED (GLCNAC. .)	Qy	944	PPCTIDVYIMVKWMDSECRPRFRELSEFSRMDRDPORFVVIQNEH-LGPAASPLDS	1002
FT	CARBOHYD	559	577	N-LINKED (GLCNAC. .)	Db	942	PPCTIDVYIMVKWMDSECRPRFRELSEFSRMDRDPORFVVIQNEH-LGPAASPLDS	1001
FT	CARBOHYD	580	589	N-LINKED (GLCNAC. .)	Qy	1003	TFYRSLLEDMDGDLVDAEYLVPOQGFCCDP-----APGA	1039
FT	CARBOHYD	593	614	N-LINKED (GLCNAC. .)	Db	1002	KFFQNLLEDLEDMDMAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSPPAYTPMS	1060
FT	CARBOHYD	617	625	N-LINKED (GLCNAC. .)	Qy	1040	GMVYHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEGAGSDVFGDGLGMAAGLQ	1099
FT	CARBOHYD	625	633	N-LINKED (GLCNAC. .)	Db	1061	GSQFVYQDGGFATQOG---MPMPYATATSTIPEAPVA--QCATAEFDDSCNGTLRKPV	1115
FT	CARBOHYD	632	633	N-LINKED (GLCNAC. .)	Qy	1100	SLPHDPSLPQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVNPQVDRPQPPSP	1152
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Db	1116	VPHVEDSSSTQRYSDPTVPAPERPNRAELDEEGYMTPMHDKPQOYLNPNVE-----	1167
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Qy	1153	REGPLPAARAGATLERAKTILSPKGVVXDVFAFGGAVENPEYLTQGGGAAPQHPPPA	1212
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Db	1168	ENPFVSR-----KNGDLQ-----ALDNPEYHSSAG-----PPKA	1198
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Qy	1213	-----FSPAFDNLVYVWDDPPPERGA--PPST	1236
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Db	1199	EDEVNEPLYNTFTNALGNAEYKMSLLSVPEKAKKAFNDPDYWNHSLPPRSTLQHPDY	1258
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Qy	1237	FKGTPT-----AENPEYL	1249
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Db	1259	LOEYSTKYFYKQNGRIRPIVAENPEYL	1285
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Qy	1297	TACPYNYLSDVGSCTLVCPHLNVEVTAEDGTQRCCKSPKPCARVCYGLGHEHLREVRV	356
FT	CARBOHYD	633	633	N-LINKED (GLCNAC. .)	Db	291	KKCPHFNV-VDSSECVRACFPSSKNEV--EENGKMKCKECTDICPKACDGI GTGSLMSAQTV	348

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;

Query Match 43.8%; Score 2984; DB 1; Length 1308;
Best Local Similarity 45.4%; Pred. No. 7.6e-153;
Matches 612; Conservative 192; Mismatches 383; Indels 160; Gaps 29;

Qy	1	MELA-ALCRWGLL--ALLPFGAASQTQCTDMKRLPLASPETHLDMLRHLHYQGCQVQ	57
Db	1	MKLATGLWVGSLLVAARTVQPSASQSVCACTENKLSLSLEQOYRALRKYENCVVM	60
Qy	58	GNLELTYLPTNASLFLQDTEOVGYVLIHNOVQVPLQRLRIVRGTLFEDNYALVL	117
Db	61	GNLEITSIEHNRDLISFLRSREVGYVVALNQFRLPLENRLIRGTLKLYEDRYALAIF	120
Qy	118	DNGDPLNNQVYKANSKFGITELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNN	177
Db	121	LN-----YRKDGN--FGQLGLKNTLILNGGVVDQNKFLCYADTIHWQDIVRNPW	171
Qy	178	QLALTLDITNRSRACHPCSPMKSGRCWGESSEDCQSLTRITVCAGGC-ARCKGPLPTDCC	236
Db	172	PSNMTLVSTIGSSGCRGCHKSCG-CRWGPTENHCQTLLTRTVCAEQCDGRCYGFVSDCC	230
Qy	237	HEOCAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCV	296
Db	231	HRECAAGCGPKDTCDFACMNFNDSGACVTCQPTFYNNPTTFOLEHFNKAYTYGAFV	290
Qy	297	TACPYNYLSDVGSCTLVCPHLNVEVTAEDGTQRCCKSPKPCARVCYGLGHEHLREVRV	356
Db	291	KKCPHFNV-VDSSECVRACFPSSKNEV--EENGKMKCKECTDICPKACDGI GTGSLMSAQTV	348

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (BC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueker W., Raulf F.,
 RA Telling A., Robertson S.M., Scharlt M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharlt M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright.. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X16891; CAA34770.2; -;
 DR PIR; S06142; S06142.
 DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 643 665 POTENTIAL.
 FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 710 977 PROTEIN KINASE.
 FT NP_BIND 716 724 ATP (BY SIMILARITY).
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 226 BY SIMILARITY.

FT	DISULFID	237	245	BY SIMILARITY.
FT	DISULFID	241	253	BY SIMILARITY.
FT	DISULFID	256	265	BY SIMILARITY.
FT	DISULFID	269	296	BY SIMILARITY.
FT	DISULFID	300	311	BY SIMILARITY.
FT	DISULFID	315	330	BY SIMILARITY.
FT	DISULFID	333	337	BY SIMILARITY.
FT	DISULFID	504	513	BY SIMILARITY.
FT	DISULFID	508	521	BY SIMILARITY.
FT	DISULFID	524	533	BY SIMILARITY.
FT	DISULFID	537	553	BY SIMILARITY.
FT	DISULFID	556	569	BY SIMILARITY.
FT	DISULFID	560	577	BY SIMILARITY.
FT	DISULFID	593	615	BY SIMILARITY.
FT	DISULFID	618	626	BY SIMILARITY.
FT	DISULFID	622	634	BY SIMILARITY.
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	144	144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	501	501	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	621	621	N-LINKED (GLCNAC. .) (POTENTIAL).
SEQ	SEQUENCE	1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;		

Query Match 39.9%; Score 2715.5; DB 1; Length 1167;
 Best Local Similarity 45.4%; Pred. No. 1.7e-138;
 Matches 577; Conservative 162; Mismatches 387; Indels 145; Gaps 27;

QY	4	AALCRWGLLLALLPPGAAST---	OVCTGTDMKRLRASPETHLDMLRHLHYGCGVVOGN	59
DB	8	AALQ--LLLVLSTSRCCSTDPDRKVCQGTNSQMTM---	LDNHYLKKMKYSGCNVLEN	62
QY	60	LELYLPTNASLSFLQDIQEVGYVLI	AHNOVRQVPLQRLRIVRGTLQFEDNYALVDN	119
DB	63	LEITYTQENQDLSFLQSIQEVGYVLI	AMNEVSTIPLNLRILRQNLVEGNTLLVMSN	122
QY	120	GDPLNNQYKANSK----	FIGITELQRLSLTEILKGGVLIQRLNPOLCYODTILMKDIFHK	175
DB	123	-----YQKNPSPDVYQVGLKQLQSLNLT	TEILSGGVKVSHPNLLCNVETINWMDIVDK	175
QY	176	NNQALTLIDTNRSRACHPCSPMCKSGRCWSESSED	CCSLTRTVCAAGC--ARCKPLPTD	234
DB	176	TSNPTMNLIPHAFCQCKDHGCVNGSCWAPGPGHCQK	FTKLLCAECNRCRGPKPID	235
QY	235	CCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCP	ALVYNTDTFESMPNPEGRYTFGAS	294
DB	236	CCNEHCAGGCTGPRATDCLACRDFNDGCTCKDTCPP	KIYDIVSHQVVDNPNIKYTFGAA	295
QY	295	CVTACPNYLSLDVSGSCTLVCLPLHNOEVTADGTQ	RCSEKSKPCARVCYGLQMEHLREVR	354
DB	296	CVKPCPSNVVYTE--GACVRS	CSAGMLEVD--ENGRSKPCDGVCPKCDGIGIGLSNTI	353
QY	355	AVTSANTIQEPAGCKKIFGLSFLPESFDGDPASNTA	PLQPEQLQVFTLEETGYLYSA	414
DB	354	AVNSTNRSFNCYKINGDIILNRNSFEGDPHYKIGT	MDPEHLMNLTTVKEITGYLYVMW	413
QY	415	WPSDLPDLSVFQNLQVIRGRILNHGAYS--LTQGLGI	SWLGLRSRLBELSGSLALHNHTH	473
DB	414	WPNNTSLSVFQNLLEIRGRITTFSGRFSVVVVQVR	HLQWLGLRSLKEYSAGNVILKNTLQ	473
QY	474	LCFVHTVPDQLFRNPQALLHTANRPEDECVGSLACH	QICARGHCNGPGTQVCNCQ	533
DB	474	LYANTINWRRLFRSEDSQSIYDART-----ENQTC	NNECEDGCGPGPTMCVSLH	526
QY	534	FLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQ	QNGSVTCFGEPAQCACVAKXDP	593
DB	527	VDRGRCVASCNLLQGLGPREAQVDGRCVQCHOECL	VQTDLSLTCYGPGPANCSKAHQD	586
QY	594	PFCVARCPGKVPDLSYMPIWKFPDEEGACQPCP	INCTHSCVDSCLDDKCPAEQASPLTS	653

```
Db 587 POCIPRPHGILGDGDTL-1WKVADKMGQCQPCQHCNCTGCGSPGLSGCRGD-IVSHSSL 644
Qy 654 IVSAVGVILLVVLGVVFGILIKRROOKIRKYMRRLLQETELVEPLTPSGAMPNQAMR 713
Db 645 AVGLVSGLLTIVALLIVVLLRRRIK-RRTIRCLLQELVEPLTPSGQAPNQAFLR 703
Qy 714 ILKETLRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYV 773
Db 704 ILKETEFKDRVLGSGAGFTVYKGLWNPDCENIRIPVAIKVIREATSPKVNQVLEAYV 763
Qy 774 MAGVGSPVSRLLIGLICLTSTVQLTQMPYGCCLLDHVRNRLGSDLLNWCWQIAKGM 833
Db 764 MASVDHPHVRLGICLTSAVQLTQMPYGCCLLDVYRQHERICGWLNNWCQIAKGM 823
Qy 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEVHADGGKVPKKNWALE 893
Db 824 NYLEERHLVHRDLAARNVLKSNHVKITDGLSKLTADEKEYQADGGKVPKKNWALE 883
Qy 894 ILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
Db 884 ILQWTVTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVYM 943
Qy 954 IMVKCWMIDSECPRELVSERMRARDQRFVWIONEDLGPASPLDSTFYRSLLEDD 1013
Db 944 IILKCMIDSPSRPRELVGEFSQWADPSRYLVIQG---NLPSSLDRRLFSRLSSDD 1000
Qy 1014 MGDVDAEYLVPOQGFCCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPS 1073
Db 1001 --DVDADEYLLPKRI-----NRQGS-----E 1021
Qy 1074 PLAPSEGAGSDVFDGOLMGAAKGLSLTHDPSPLQRYSEDPTV-PLPSETDGYVAPLT 1132
Db 1022 PCIPPTGH-----PVRENSITLRNLSDTQNALEKLDGH----- 1056
Qy 1133 CSQPQEVNPDVVRPQ-----PSPRE-----GPLP-AARPAGATLERAKTSLPGKNG 1179
Db 1057 -----EYVNPQGETSRSLSDIYNPNYEDLTDGMPVSLSQEATNFSRPEYLNQNS 1111
Qy 1180 VVKDVFAGGAVENPEYLTPOGGAAPQHPHPPAFSDNLYYWDQDPPRGAPPSTFKG 1239
Db 1112 L----PLVSSGSMDDPY---QAG-----YQAFA-----LPQTGALTNGMF 1146
Qy 1240 TPTAENPEYLG 1250
Db 1147 LPAENLEYLG 1157

RESULT 9
ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (BC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90083234; PubMed=2687875;
RX Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MTM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
FT POTENTIAL
FT RECEPTOR_PROTEIN_TYROSINE_KINASE_ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN_KINASE.
FT ATP (BY SIMILARITY).
FT ATP (BY SIMILARITY).
FT BY SIMILARITY.
FT BY SIMILARITY.
```


RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=9609635; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Kolland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
recombinant protein";
RL Gene 165:279-284 (1995).
RN [2]
RP REVISIONS TO 85; 513 AND 565.
RA Hellyer N.J., Kolland J.G.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuroregulins and their putative receptors, ErbB2 and
ErbB3, is induced during Wallerian degeneration";
RL J. Neurosci. 17:1642-1659 (1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U29339; AAC28498.2; -;
DR EMBL; U52530; AAC53050.1; -;
DR HSPB; P11362; LFGLK
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 19
FT CHAIN 20 1339
FT DOMAIN 20 643
FT DOMAIN 644 662
FT TRANSMEM 663 1339
FT DOMAIN 183 259
FT DOMAIN 707 964
FT NP_BIND 713 721
FT BINDING 740 740
FT ACT_SITE 832 832
FT DISULFID 186 194
FT DISULFID 190 202
FT DISULFID 210 218
FT DISULFID 214 226
FT DISULFID 227 235
BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1028 1028 L > P (IN REF. 3).
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;
Query Match 34.7%; Score 2359.5; DB 1; Length 1339;
Best Local Similarity 40.8%; Pred. No. 2.6e-119;
Matches 523; Conservative 172; Mismatches 433; Indels 155; Gaps 34;
QY 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMLKRLPASPTHLDMLRLHLYQGCQVQGN 59
DB 7 LQVLC-----FLSLARGSEMGNSQAVCPGTGLNGSLVTGDNADNQYQTLKYKECEVWGN 62
QY 60 LEITYLPTNASLFLQDIQEVQGYVLIHNNQVQPLQRLRIVRGTLQFEDNYALAVLDN 119
DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSVLPLNLRVVRGTQYDGKFAIFVNLN 122
QY 120 GDPLNNQYIKANSKFITGLTELQRLSITELKGGVLTORNPOLCYQDTILWKDIFHKNQL 179
DB 123 -----YNTNSSH-ALRQLKFTQLTEILSGGVYIEKNDKLCMDITDRVVR--- 170
QY 180 ALTLIDTNRSRACHPCSPCKSRCKGSESSDCQSLTRTYVCAGC-ARCKGPLPTDCCHE 238
DB 171 GAEIVVKNNGANGCPPCHEVCKG-RCMGPGDDCQIILTKTICAPCNCRCGFPNPQCCHD 229
QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTA 298
DB 230 ECAGGCGSPQDTCFACRRFNDSGACVPRCPPELVVYNKLTFOLEPNPHTKYQYGGVCVAS 289
QY 299 CPVNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMHLEHREAVTS 358
DB 290 CPNFFV-VDQTFVCRACPPDKMEVD-KHGLKMCPCGGLCPKACEGTGSG--SRYQTVD 345
QY 359 ANIQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEITGYLVISAMPDS 418
DB 346 SNIDGFVNCTKILGNLDFLITGLNVDPWHKI PALDPEKLVNFTVREITGYLNIQSWPPH 405
QY 419 LPDLSVFQNLQVIRGRIHNGAYS-LTLOGLGISWLGRSLRSLRGSLALIHHTHLCFV 477
DB 406 MHNFSVFSNLTTIGRSLYNRGFSLLMKNLNVTSLGFRSLKEISAGRVVISANQQLCYH 465
QY 478 HTVPWDLPRNPHQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPGTQCNCVCSFLR 536
DB 466 HSLNMTLLRGLGPSEERLDIKYDRPLGECLEAGRVKVCDFPLCSSGGCGWGPQGLSCRYNR 525
QY 537 GQECVEECRVLOGLPREYVNRHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFPC 596
DB 526 EGVCVTHCNFLOQEPREFVHEAQCFSCHPECLMEGTSTCNGSGGSACARCAHFRDGPCH 585
QY 597 VARCPGSKVPDLSYMPDWKFPDEEGACQPCPINCTHSC--VDLDDKGCFAEQRASPLTSI 654

```
Db 586 VNSCPHGILG--AKGPIYKPDQNECRPCHENCTQGCNGPBLQDCLGQAEVLMKPHLV 643
Qy 655 VSAAVGIILVVGVVGLIKIRKQKIR-KYTMRRLLQETELVPLTPSGAMPNQAOVR 713
Db 644 IAVTVG--LAVILMLGGFLYWRGRIQNKRAMRYLGRGESSIEPLDPS-EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAGTGVYKGIWIDGENVKIPVAIKVURENTSPKANKEILDEAVY 773
Db 701 IFKETELRKVLKVLGSGVGTGVHKGWIWIPGESIKIPVCIKVIEDKSGRQSFQAVDHLMA 760
Qy 774 MAGVGSPPVRLRLGLTSTVOLVTQMLPYGCLLDHVNRNRLGSGQDLNMCWQIAKGM 833
Db 761 VCSLDHAHIVRLGLGCPSSQLQVLTQYPLGSLLDHVHQHRTLPQLLLNMGVQIAKGM 820
Qy 834 SYLEDVRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGKGVPIKMALES 893
Db 821 YLLEHSMVHRDLALRNVLKSPSQVQVADFVADLLPPDDKQLLHSEAKTPIKMALES 880
Qy 894 ILRRFTHOSDWSYGVTVWELMTGCAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYM 953
Db 881 IHFGKYTHOSDWSYGVTVWELMTGCAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYM 940
Qy 954 IMVKCWMIDSECRPFRELVSFMRMDPQRFVVIQNEIDLGPASPLDSTFYRSLDEDD 1013
Db 941 VMVKCWMIDENIRPTFKELANEFTMRADPPRYLVIKRAS-GPGTP--PAAPSPVLTNKE 997
Qy 1014 MGLDVAEYLVPOQGFCCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSE- 1068
Db 998 L-----QEALELEP-----DLDLDLEAEESGLATS 1023
Qy 1069 -----EAPRSLAPSEG-----AGSDVFDGLGMGAAGKQLSLPHD 1105
Db 1024 LGSALSUPTGLTRPRGQSLLSPSSGYMPNMQSSGLGACLDLSAVLGGREGQFSRPSISLH- 1082
Qy 1106 PSPLQRYSEDPTPLPSTDGVV-----APL-----TC-----SPQPE-----VYNQPDV 1145
Db 1083 PIPGR-----PASESEGHVTGSEAELEKQVCSRGRSRSPRGRGSAYHSQRHS 1135
Qy 1146 RPQPPSPREGP-----LPAARPAGATLERAKTLSP-KGNQGVV-----KDVPAF 1187
Db 1136 LTPVTLPSPLGLEEDGNGVMPDTHLARGASSREGTSSVGLSSVLGTEEDED----- 1191
Qy 1198 GGAIVENPEYLPQGGAAPOPHPP 1210
Db 1192 ----EYEYMNMRKRGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT: 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RL subdomains of the receptor protein.";
RN Genetics 137:531-550(1994).
RN (2)
RP REVISIONS.
```

```
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBSJ databases.
RN (3)
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RL hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN (4)
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RL Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN (5)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RL specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yeaman M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazewicz R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Buesam D.A., Butler H., Cadenhead E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2000).
RN (7)
RP SEQUENCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
```

```

RT growth factor receptor.";
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/fib alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Perrimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor.";
RL Cell 89:13-16(1997).
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF052754; AAC08536.1;
DR EMBL; AF052753; AAC08536.1; JOINED.
DR EMBL; AF052754; AAC08535.1;
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1;
DR EMBL; K03417; AAA51460.1;
DR EMBL; K03416; AAA50965.1;
DR EMBL; K03418; AAA51461.1;
DR EMBL; AF109077; AAD26134.1;
DR EMBL; AF109078; AAD26132.1;
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1;
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1;
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1;
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1;
DR EMBL; AE003454; AAF46732.1;
DR EMBL; X02293; CAA26157.1;
DR EMBL; X78920; CAA55523.1;
DR EMBL; X78918; CAA55521.1;

```

```

EMBL; X78919; CAA55522.1;
DR PIR; A00640; GQFFE.
DR HSSP; P11362; LFQK.
DR FlyBase; FBgn0003731; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
KW Developmental protein.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 869 889 POTENTIAL.
FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 938 1198 PROTEIN KINASE.
FT NP_BIND 944 952 ATP (BY SIMILARITY).
FT BINDING 971 971 ATP (BY SIMILARITY).
FT ACT_SITE 1063 1063 BY SIMILARITY.
FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).

Query Match 29.0%; Score 1976; DB 1; Length 1426;
Best Local Similarity 33.1%; Pred. No. 1.le-98;
Matches 471; Conservative 184; Mismatches 431; Indels 338; Gaps 42;

QY 24 QVCTGDMKRLRSPETHLDMLRHLVQGVQGNLELTLYPT-NASLSFLQDIQEVQ 82
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
100 KICIGTKSLRSLVPSNKEHRYNLRDRYNTCTYDGNLKLTLNPNLNLDSFLDNIREVTG 159
QY 83 YVLIAINQVRQVPLQRLIRVGTQLF-----EDNYALAVLDNGDPLNNQVIKANSKIFGI 137
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 YLISHVDVKKVVPFKLQIRGTLFSLSEVEEKALFV-----TYSK-----M 203
QY 138 TELQSLASLEILKGGVLIORNPOLCYODITLWKDIFHNQNLALTLDITNRSRACHPCSP 197
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 YLLEIPDLRDLVNGQGVFNHNNYLNCHMRTIQSEIVSNGTDAYNYVDFTAPERCPKCHE 263
QY 198 MCKGSRCSWGESSEDCQSLRTRTVCGAGCA--RCKGPLPTDCCHQCAGCTGPKHSDCLAC 255
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 SCTHG-CWGECPKNCQFKSLTCSPOCAGRCYGPKECHLFCAGGCTGTOKDCIAC 322
QY 256 LHPNHSIGICELHCPALVTYNTDTFESMPNPEGRYTGASCVTACPNYNYLSTDVGSCTLVC 315
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 KNFFDEAVSKEECPMRKYNPTTYVLETNPEGRYAYGATCVKECP-GHLLRDNGACVRS 381
QY 316 PLHNOEVTADGTQRCCKSKPCARVCYGLGMEHLREAVTSAIOEFAGCKKIFGSLA 375
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
382 PQDKMDKGE-----CVPNGPCPKTCPGVTVLH-----AGNIDSFRNCTVIDGNIR 428
QY 376 FLPESPDG--DPASNTA-----PLOPEQLQVFETLEEITGYLYISAWPDSLPLSVFQN 427
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
429 ILDTQTSFGQDVVANYTGMGPRIPLDPERREVFSTVKEITGYLNIEGTHPQFNLSYFRN 488
QY 428 LQVIRGRILHNGAY-SLTQGLGIGISWIGLRSRELQSGIALIHNNTHLCFVHTVPDQLF 486
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
489 LETIHGRQLMESMFALAIVKSSLSYLEMRNLKLISSGSVIQRNRLDLCVYSIRNPAIQ 548
QY 487 RNPHQALLTANRPEDECVGEGGLACHQLCARHCWGPPTQCVCNCSOFLRGQSCVCECRV 546
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
549 KEPEQKVWNENLRADLCENKNGTICSDQCNEDCGCWAGTDQCLCTKNFNFGTICADCGY 608
QY 547 LQGLPREYVYVNRHCLPCHPECPQNGSVTCFCGPEADQCACAHYKDPFCVARCP----- 601

```

```
Db 609 ISNAYK--FDNRKTKICHEPCR-----TCNAGADHCQECVHVRDQHCVCSECPKKNYN 660
Qy 602 -----SGVK-----PDLSTYPIW 614
Db 661 DRGVCRECHATCDGCTGPKDTIGIGACTTCNLAINDATVKRCLLKKDKCPD-GY--FW 717
Qy 615 KF--PDEGACQP-----NCTH----- 632
Db 718 EYVHPQEOGSLKPLAGRAVCRKCHPLCLCTNYGYHEQVCKSKCTHYKRRQECETCPADH 777
Qy 633 -----SC-----VOLDDKG-----CPAQ 646
Db 778 YTBQRECFORHPECNCTGFGADDCKSCRFKFDANETGPPYVNTMFNCTSKCPLEM 837
Qy 647 R-----ASPLTS-----IVSAVVGILLVVVLGVFGILLIKRRQ 679
Db 838 RHVNYQYTAIGPYCAASPPRSKITANLDVNMIFIITGAVLVPTICILCV--TYICRQK 895
Qy 680 QKIRKYT--MRLLOETELVEPLTPSGAMPNOAQRILKTELKRVKVLGSGAFGVTVYKG 737
Db 896 QKAKKETVMTMALSGCSDSEPLRPSNITCANLCKLRIIVKDAELRKGGLMGAFGRVYKG 955
Qy 738 IWIPOGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYVSRLLGLICTSTVOLV 797
Db 956 VWPVEGENVKIPVAIKELKSTGAESSEFEIYIMASEBHVNLKLLAVCMSSQMMLI 1015
Qy 798 TOLMPYGCLLDHVRENRLGSGODLLNMCIOAKGMSYLEVDRLVHRDLAARNVLKSPN 857
Db 1016 TOLMPLGCLLDYVNRNRDKISGKALLNWSIOAKGMSYLEKRLVHRDLAARNVLQTPS 1075
Qy 858 HVKITDFGLARLLDIDETEHADGGKVPKPKMALESILRRRFTHQSDVMSYGVTVWELMT 917
Db 1076 LVKITDFGLAKLLSSDSNEYKAAGSKMPIKWLALCIRNRVFTSKSDVWAFGVTIWELLT 1135
Qy 918 FGAKPYDGIPIAREIPDLLEKERLPOPPDICTIDVYIMVWCMIDSECRPRFRELVSFS 977
Db 1136 FGQPHENIPAKDIPDLIEVLGKUEQPEICSLDIYCTLGLSCWHLDAAMRPTFKQLTTVFA 1195
Qy 978 RMARDPORFVVIQNEGLG--PASPLDSTFYRSLLEDD---DMGDLVDAAEYLVPOQGF 1032
Db 1196 EFARDPGYLAIPGDKFRLPA-----YTSQDEKILIRKLAPTDTGSEAIARPDYLIQ 1248
Qy 1033 PDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEEAPEAP-----RSLAPSEAGSDV 1086
Db 1249 PKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPSNKNSSTGDDER 1287
Qy 1087 DG---DLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYVNP 1143
Db 1288 DSSAREVGVGNLR-----LDLPVDEDDYLMPTCOQPGPNNNNM 1325
Qy 1144 DVPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYL---TP 1199
Db 1326 N-----NPNQNNAAVGAAGYM-----DLIGVPSVDNPEYLLNAQTL 1364
Qy 1200 QGGAAPQH-----PPAFSP-AFDNLYWD 1224
Db 1365 VGESPIPTOTIGIPVGGPGTMEVKVPMGPGSEPTSDHEYYND 1408
```

RESULT 12

```
ERBB ALV STANDARD; PRT; 634 AA.
ID ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilleen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFVLV.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 6e-87;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;
Qy 587 CAHYKDPFCVARGSPGVKPDLSYMPIWKFDPDEGACQPCPINCTHSCVDLDDKGPASQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCQLCHPNCTRGCKGPLEGCP--- 58
Qy 647 RASPLTSTVSAAV-GILLAVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSCA 705
Db 59 NSKTPSTAAGVGLCLVVGIGLYLRRR-HIVRRKTLRRLRQLRELVEPLTPSGE 117
Qy 706 MPNOAQRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 APNOAHLRLAKETEFKKVKVLGSGAFGTYYKGLWIPEGEKVIPIVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPYVSRLLGLICTSTVQLVTLQMPYGCLLDHVRENRLGSGODLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGLICTSTVQLITQMPYGCLLDYIREHKDNTGSGVLLNW 237
Qy 826 CMOJAKGMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQJAKGMYLEERRLLVHRDLAARNVLKTPQHVKITDFGLAKLLGADEKEYHAGGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKERLPOPP 945
```


or send an email to license@isb-sib.ch).

```
CC -----
CC EMBL; M20386; AAA48760.1; ..
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00111; PROTEIN KINASE DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.7%; Score 1614; DB 1; Length 703;
Best Local Similarity 44.3%; Pred. No. 1.3e-79;
Matches 313; Conservative 113; Mismatches 253; Indels 28; Gaps 11;

QY 8 RWGLLLALLPPGAA-----STOVCTGTDMKRLPASPETHLDMLRHLVYQGVVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKVCGQTNNKLTQLGHVEDHFTSLQRMNNCEVLSNLE 72

QY 62 LTYLPTNASISFLQDIOEVGYVLIHQAHQVRQVPLQRLRIVRGTFQDFEDNYALAVLDNGD 121
DB 73 ITVVEHNRDLTFLKTIQEVAGYVLIHQAHQVRQVPLQRLRIVRGTFQDFEDNYALAVLDNGD 132

QY 122 PLNNQYIKANSKPIGITELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQAL 181
DB 133 MNKTQ-----GLRELPMKRLSEILNGVKISNNPKLNCNMDTLVNDIIDSRL-PL 182

QY 182 TLID-TNRSRACHPCSPMKGSRGWGSSSDCQSLRTVYVAGGCA-RCKGPLPTDCCHQ 239
DB 182 TLID-TNRSRACHPCSPMKGSRGWGSSSDCQSLRTVYVAGGCA-RCKGPLPTDCCHQ 239
```

```
Db 183 TVLDPASNLSSCPKCHPNCCTEDHCWGAGEQNQCOTLTKVICAQQCSGRCRGKVPSPDCCHQ 242
QY 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFBSMPNPEGYTTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACRKRFRDDATCKDTCPLVLYNPTTYOMDVNPEKYSFGATCVREC 302
QY 300 PYNLYSTDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVATSA 359
DB 303 PHNYVTDHSGSVRSCTNTDTYEV-EENGVRKCKKCDGLCSKVCNGIGIGELKGLSINAT 361
QY 360 NIQEFACKKIFGSLAFLESFDDPASNTAPLOPEQLOVFTLEETIYLYISNWPDSL 419
DB 362 NIDSFKNCTKINGDVSLPVAFLGDAFTKTLPLDPKLDVFTYKVEISGFLLIQAWPDNA 421
QY 420 PDLISVFQNLQVIRGRILHNGAYSULTLOGLGISMLGLSLRLSLRELGLSGLALIHNHNL 479
DB 422 TDLYAFENLEIRGTQKHGOYSLVAVNLKIQSLGLSLKKEISGDIAIMKNKNLCYADT 481
QY 480 VPMQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCGWGPPTQCVNCSQFLROE 539
DB 482 MNWRSLEATOSQTKIIQNRKNKNDCTADRHVCDPLCSVDGCGWGPFPCHFCRFSRQKE 541
QY 540 CYEECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFQPEADQCVACAHYKDPFPC 596
DB 542 CVKQCNTLQGEPRERFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VARCPGKVPDLSYMPWIKFPDEBGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVS 656
DB 602 VKACPAGVLGENDTL-VKKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTPSIAA 657
QY 657 AVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLRLLQETELVEPLTP 702
DB 658 GWGGLLCLVVGVLGIGLYLRRR-HIVRKRTLRELLQERELVEPLTP 703
```

Search completed: July 22, 2003, 08:44:24
Job time : 20.2304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_arChaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6253	91.8	1259	6 O18735	O18735 canis famil
2	3156	46.4	1209	11 Q9QX70	Q9QX70 rattus norv
3	3128	45.9	1210	11 Q9EP98	Q9EP98 mus musculu
4	2749	40.4	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2729.5	40.1	1137	13 Q9WGF6	Q9WGF6 gallus gall
6	2315	34.0	1328	13 P79754	P79754 fugu rubrip
7	2047.5	30.1	1433	5 Q8B1H9	Q8B1H9 anopheles g
8	1798.5	26.4	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.5	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.3	729	15 Q86712	Q86712 avian rous-
11	1718	25.2	567	15 Q86714	Q86714 avian rous-
12	1697.5	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1653.5	24.3	962	15 Q64895	Q64895 avian eryth
14	1645	24.2	545	15 Q85468	Q85468 avian eryth
15	1519.5	22.3	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1503.5	22.1	643	11 Q9ERV6	Q9ERV6 mus musculu

17 1288 18.9 1193 5 Q9YLX8
18 1215.5 17.9 1368 5 Q23821
19 1165 17.1 1717 5 Q26566
20 1145 16.8 527 13 Q90836
21 1014.5 14.9 478 11 Q9ES02
22 961.5 14.1 599 13 Q9PSH2
23 906 13.3 165 4 Q14256
24 887 13.0 176 11 Q923V5
25 806.5 11.8 346 13 P11776
26 778 11.4 435 5 Q8SZW1
27 758.5 11.1 1362 13 Q9PV24
28 754.5 11.1 311 13 Q9162
29 729 10.7 1671 5 Q9NJV5
30 724 10.6 331 4 Q9BUD7
31 723 10.6 149 6 Q9BG66
32 703.5 10.3 1368 13 Q8UW85
33 697 10.2 1418 13 Q93457
34 683.5 10.0 1369 13 Q8UW86
35 674 9.9 1358 13 Q73798
36 669.5 9.8 1472 5 Q9U5A8
37 657 9.6 1412 13 Q8UW84
38 643.5 9.5 1418 13 Q8UW83
39 640 9.4 1245 13 Q9YGH8
40 640 9.4 1371 11 Q9QVW4
41 621.5 9.1 2144 5 Q9VD94
42 598 8.8 987 11 Q91YM0
43 595 8.7 935 4 Q96L35
44 595 8.7 987 11 Q99MR2
45 587.5 8.6 1036 4 Q07912

ALIGNMENTS

RESULT 1
O18735 PRELIMINARY; PRT; 1259 AA.
ID O18735
AC O18735;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erba-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erba-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; Ylp_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		91.8%; Score 6253; DB 6; Length 1259;
Best Local Similarity		91.6%; Pred. No. 0;
Matches 1154; Conservative 40; Mismatches 60; Indels 6; Gaps 2;		
Qy	1	MELAAALCRWGLLLALLPGCAASTOVCTGTDMLRLPASPTHLDMLRHLYQCQVVOGNL 60
Db	1	MELAAWCRWGLLLALLPGSAGTQVCTGTDMLRLPASPTHLDMLRHLYQCQVVOGNL 60
Qy	61	ELTYLPTNASLFLQDIOEVQGYVLIHNVQVPLQRLRIVRGTQLFEDNYALAVLNG 120
Db	61	ELTYLPANASLFLQDIOEVQGYVLIHNSQVRIPLQRLRIVRGTQLFEDNYALAVLNG 120
Qy	121	DPLNNQYIKANSKFTIGTELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKQNQLA 180
Db	121	DPLEGGIPAPGAACQGLRELQRLSUTEILKGVLIQRSPLCHQDTILWKDIFHKQNQLA 180
Qy	181	LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGCCARCKGPLPTDCCHEQC 240
Db	181	LTLIDTNRFSACPSPCKADACHWAGSDCQSLTRTVACGCCARCKGPQPTDCCHEQC 240
Qy	241	AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db	241	AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300
Qy	301	YNLSTDVGSCTLVCLPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db	301	YNLSTDVGSCTLVCLPHNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOVFTLEBITGYLISAWPDSL 420
Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELRFVEALEBITGYLISAWPDSL 420
Qy	421	DLVSFQNIQVIRILHNGAYSLTLOGIGISWGLRSLRELGSGLALIHNNTHLCFVHTV 480
Db	421	NLSVFONLRVIRGVLDGAYSLTLOGIGISWGLRSLRELGSGLALIHNRARLCFVHTV 480
Qy	481	PDQLFRPHQALLHTANRDEECVGEGLACHQLCARGHCWGPPTQCVNCSQFIRGQEC 540
Db	481	PDQLFRPHQALLHSANRDEECVGEGLACYP-CAHGHCGWGPPTQCVNCSQFIRGQEC 539
Qy	541	VEECRVLOGLPREYVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db	540	VEECRVLOGLPREYVYKDRYCLPCHSECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 599
Qy	601	PSGVKPDLSYMPIWKFPEEGACQPCPINTHSCVDLDDKGPAPQASPLTISIVSAVVG 660
Db	600	PSGVKPDLSFPMKFADEEGTCQPCPINTHSCADLDEKGPAPQASPLTISIIAAVVG 659
Qy	661	ILLVVVLGVFGILIKRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
Db	660	ILLAVVVLGVFGILIKRRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 719
Qy	721	RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db	720	RKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 779
Qy	781	YVSRLLGICLSTVOLVTQMPYGCCLDHVRENRLGSLQDLNMCQIAKMSYLEVD 840
Db	780	YVSRLLGICLSTVOLVTQMPYGCCLDHVREHRLGSLQDLNMCQIAKMSYLEVD 839
Qy	841	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEXHADGKVPDKWMALESILRRFT 900
Db	840	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEXHADGKVPDKWMALESILRRFT 899
Qy	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960
Db	900	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKWM 959
Qy	961	IDSECRPRFRELVSFESMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1020
Db	960	IDSECRPRFRELVAEFESMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGDLVDA 1019
Qy	1021	EYLVPQGGFFCPDAPACGAGWVHRHRSSTRSGGDLTLGLEPSEEAERPSLAPSEG 1080
Db	1020	EYLVPQGGFFCPEPTGAGGTAAHRRHRSSTRNGGDLTLGLEPSEEBEPKSLAPSEG 1079
Qy	1081	AGSDVDFDGLGGAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1140
Db	1080	AGSDVDFDGLGGAAGLQSLPHTDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1139
Qy	1141	NQDVRPQPPSPREGPLPAARPAAGATLER-----AKTILSPGKNGVVKDVFAGGAVENPE 1195
Db	1140	NQPEWVPQPLALEGLPSPRPAAGATLERPKTILSPKTLSPGKNGVVKDVFAGGAVENPE 1199
Qy	1196	YLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPSPSTPGTPTAENPEYLGLDVVP 1255
Db	1200	YLAPRGAAPQHPHPPAFSPAFDNLYYWDQDPSERGSPPSTPEGTPTAENPEYLGLDVVP 1259
RESULT 2		
Q9QX70		PRELIMINARY; PRT; 1209 AA.
ID	Q9QX70	
AC	Q9QX70;	
DT	01-MAY-2000 (Tremblrel. 13, Created)	
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)	
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=9025888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue.";	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; -;	
DR	HSSP; P11362; 1FGK.	
DR	InterPro; IPR000494; EGFR L domain.	
DR	InterPro; IPR000719; Euk pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00111; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CCLB7773 CRC64;	
Query Match		46.4%; Score 3156; DB 11; Length 1209;
Best Local Similarity		50.0%; Pred. No. 3.2e-229;
Matches 639; Conservative 168; Mismatches 356; Indels 114; Gaps 26;		
Qy	3	LAALCRWGLLLALLPPGA-ASTQVCTGTDMLRLPASPTHLDMLRHLYQCQVVOGNL 61

[illegible]

SQ SEQUENCE 1165 AA; 129614 MW; 7F7BE38D8771A74E CRC64;

Query Match 40.4%; Score 2749; DB 13; Length 1165;
Best Local Similarity 45.8%; Pred. No. 1.7e-198;
Matches 584; Conservative 162; Mismatches 382; Indels 148; Gaps 29;

QY 1 MELAAALCGLLALLPPG-AAST---QVCTGTDMLKRLPASPETHLDMLRHLYGCOV 55
DB 4 LELLE---LULLLSIGRCSTDPDRKVCQTSNQMTM---LDNHYLKMKGKMGSCNV 56
QY 56 VOGNLELYLPTNASLFLQDIQEVGYVLIHAHQVROPLORLRIVRGQTQLFEDNYALA 115
DB 57 VLENLEITYTOENQDLSFLOSIOEVGYVLIAMNEVSTIPLVNLALIRQNLQYEGNFTLL 116
QY 116 VLDNGDPLNNQYIKANSK----FIGITELQLSLTEILKGVLIOBNPOLCYQDTILMKD 171
DB 117 VMSN-----YQKNPSSPDVYQVGLKQLQNLTEILSGGVKVSHPNLLCNVETINWMD 169
QY 172 IFHKHQLALTLIDNRSACHPCSPMKGSCWGESSEDCQSLRTVTCAGCG-ARCKGP 230
DB 170 IVDKTSNPTMNLIPHAERQCKDPCGVNGSCWAPGPHCKFTKLLCAEQNRCRGP 229
QY 231 LPTDCHEOCAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTPESMPNPEGRYT 290
DB 230 KPIDCNEHCAGCTGPRATDCLACRDFNDGTCKDTCPPKIYDIVSHQVVDNPNIKYT 289
QY 291 FGASCVTACPNYLSTDVGSCTLVCLPHNQEVTAEDGTORCKSKPCARVCYGLGMEHL 350
DB 290 FGAACVKECPNSVYVTE-GACVRSACAGMLEVD-ENGKRSCKPCDGVCPKVDGIGSL 347
QY 351 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDGPASNTAPLOPQLOVFLEERTGYL 410
DB 348 SNTIAVNSNIGSFNCTKINGDIIILNRNSFGDPHYKIGPMDPEHMLNLTITVKEITGYL 407
QY 411 YISAWPDSLPDLSVFQNLQVIRGRILHNGAYS-LTLQGLGISWLGSLRSLRGSLALH 469
DB 408 VIMWMPENMTLSVQNLLEIRGRITFSRFGFVVVQVSHLQWLGSLLSKEYSAGNVILK 467
QY 470 HNTHLCFVHTVPWDQLFRNPQHALLHTANRDECEVGEGLACHQIACRGHCWGPQTQCV 529
DB 468 NTPQLRYASTINWRRLFRSEDOSIEYDART-----ENQTCNNECEDGCGWPGPTMCV 520
QY 530 NCSQFLRGQECVEECRLVQLPREYVNRHCLPCHPECOPONGSVTCFPGPEADQVACAH 589
DB 521 SCHVDRGRCVASCNLLQEPREAVQDRCVCHQECVLQVDTSLTYCGPGPANCSCAH 580
QY 590 YKDPFPCVACRSPGVKPDLSYMPIWKFDPDEGACQPCINCTHSCVLDLDDKCPAEORAS 649
DB 581 FQDGPQCIPRCPHGMGLGDGTL-INKYADKMGCQPCQCHQNCQCGSGPGLSGCRGD-IVS 638
QY 650 PUTSIVSAVVGILLVVVLGVVFGIILKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709
DB 639 HSSLAVGLVSGLLITIVALLIWLRLRRRIK-RKRTIRLLQEKELVEPLTPSGQAPNQ 697
QY 710 AQMRILKETELRKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILD 769
DB 698 AFLRLIKETEFKDRVLGSGAGTVYKGLWNPDPENIRIPVAIKVLRNTSPKANKEILD 757
QY 770 EAYVMAGVSPYVSRLLGICLTSTVQLVTQLMPYCGCLDPHVRNRRGLQSDQLLNWCMQI 829
DB 758 EAYVMASVDHPHVCRLGICLTSAVOLVTQLMPYCGCLLDYVHQHQRICGQWLLNWCQI 817
QY 830 AKGMSYLEDRVLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHADGCKVPIKWM 889
DB 818 AKGMNYLEERHLVHRDLAARNVLLKPNHVKITDFGLSKLLTADEKEYQAHGCKVPIKWM 877
QY 890 ALESLRRFRTHQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTI 949
DB 878 ALESLQWYTHQSDVWSYGVTWELMTFGSKPYDGIIPAKEIASVLENGERLPQPPICTI 937
QY 950 DVTYIMVKWIMIDSECRPRFELVSEFRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLL 1009
DB 938 EYVMIILKWMIDPSSRRPRFELVGEFSGMARDPSRYLVIQ---NLPSPSDRRLFSRLL 994

QY 1010 EDDDDGMDLVDABEYLVPOQGFCCPDAPGAGGMVHRHSSSTRSGGDLTLGLERPEEE 1069
DB 995 SSDD--DVDADEYLL-----RYKRIN-RQS----- 1018
QY 1070 APRSLAPSEAGSDVFDGDLGMAKGLQSLPHTDPSPLORYSEDPV-PLPSETDGYV 1128
DB 1019 ---EPCIPNGH-----PVRENSIALRYISDPTQNALEKDLDDGH- 1054
QY 1129 APLTCSPOPEYVNVQDVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLSP 1175
DB 1055 -----EYVNPQSETSSRLSDIYNPNYEDLTDGWPVLSLSQEAETNFSRPEVLNT 1105
QY 1176 GKNGVVDVFAFGGAVENPEYLTPOGGAAPQHPHPPAFPAFDNLYWDDPPPERGAPPS 1235
DB 1106 NQNSL---PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTG 1140
QY 1236 TFKGTPTAENPEYLG 1251
DB 1141 NGMFLPAAENLEYLG 1156

RESULT 5
Q9W6F6 PRELIMINARY; PRT: 1137 AA.

AC Q9W6F6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERB84.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=HINDBRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "Distribution of neurotrophin-1 (nrg1) and erbB4 transcripts in
RT embryonic chick hindbrain."
RL Mol. Cell. Neurosci. 13:237-258(1999).
DR EMBL; AF121963; AAD31764.1; -;
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU_3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW Kinase; Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 1137 AA; 12927 MW; 4D616436F87DC84F CRC64;

Query Match 40.1%; Score 2729.5; DB 13; Length 1137;
Best Local Similarity 47.3%; Pred. No. 4.8e-197;
Matches 539; Conservative 169; Mismatches 354; Indels 77; Gaps 24;

QY 161 LCYQDTILWKDIFHKHQLALTLIDNRSACHPCSPMKGSCWGESSEDCQSLRTVTC 220

```
Db 3 LCFADTIHQDIVRNWASNFTLVPTNGSGCGCHKSCGTG-RCWGPTENHCQTLLTKTVC 61
Qy 221 AGGC-ARCKGPTDCCHCCAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDF 279
Db 62 AEODGRCYGVYVSDCHRECAGCGPKDTCFACNFDSGACVTCQPTVYVPTTF 121
Qy 280 ESNPNPEGRYTFGASCVTACPNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCA 339
Db 122 OLEHNHNAKTYTGAFCVKCKPHFV-VDSSSCVACPCSSKMEV-EENGIMCKPCTDIPC 179
Qy 340 RVCYGLGMEHLREVRVAVTSANIOEPAGCKKIFGSLAPLPSFDGDPASNTAPLOPOLQV 399
Db 180 KACDGTGTSLSVAQTVDSDSNKDFNCTKINGNLIFLVTHGDPYHTIAAINPEKINI 239
Qy 400 FETLEBITGYLYISAWPDSLPLDLSVFQNLQVIGRILHNGAYSLLTQLGLGISMGLRSR 459
Db 240 FQVREITGYLNTQSPENWTDPRFVSNLVTIGRALYSGLSLILKQOIGITSLQFSLK 299
Qy 460 ELGSGLALIHNTHLCHFVHTVPMQDLFRNPHQALLHTANRPEDECGEGGLACHQLCARGH 519
Db 300 QISAGNIYITDNSLNCYVHTVNTSLFSTPSQKTVIHRNKAENCTADGMVNCNELCSSDG 359
Qy 520 CWGPGTQVCNCSQFLRGQCEVCECRVLQGLPREYVNAHCLPCHPECCP-QNGSVTCFG 578
Db 360 CWGPGDQCLSCKRFIRGRTCTIESCNLYDGEFREFANGSVCMEDPQCEKWMEDNMITCVG 419
Qy 579 PEADQCVACAHYKDDPPFCVACRCPGVKPDLSYMPIWKFDEEGACOPCPCINCHSCVDLD 638
Db 420 PGPDHCTKCFHFQDGNVCVEKCPDGLGANSF--IFKADDECHPCPNCTQCGRGA 477
Qy 639 DKGC-----PAEORASPLTSIVSAVV-GILLVVLGVVFGILIKRROOKIRKYT 686
Db 478 SHDCIYVPTWQSTLPOHAR-TPL--IAAGVIGGLFIIVIMGLTFAVYVRKSIK-KKRA 533
Qy 687 MRLLQETELVEPLTSGAMPNQAQMRILKETELRKVKVLGSGAFGTGYKGIWPGENV 746
Db 534 LRRFL-ETELVEPLTSGTAPNAQRLKETELRKVKVLGSGAFGTGYKGIWVPEGETV 592
Qy 747 KIPVAIKVIRENTSPKANKEILDEAYVMAGVGPVYSRLIGLICLTSTVOLVTOLMPYGL 806
Db 593 KIPVAIKILNETTGKANVEFMDALIMASMDPHLVRLLGVCLSTIQLVTLQMPHGCL 652
Qy 807 LDHVRNRRGLSGDILLNCMQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGL 866
Db 653 LDYVHEHKONIGSOLLNMCVQIAKGMVLEERLVRDLAARNVLKSPNHVKITDFGL 712
Qy 867 ARLLDIDETEHADGKVPIKMALESILRRRTHQSDVMSYGVYTWELMTFCAKPYDGI 926
Db 713 ARLEGDEKEYNADGKMPKMALEBCHYRKFTHQSVDVMSYGVYTWELMTFGKPYDGI 772
Qy 927 PAREIPDLLEKGERLPQPICTIDVYIMVIMVCMIDSECRPRELVSEFSRWARDPQRF 986
Db 773 PTREIPDLLEKGERLPQPICTIDVYIMVIMVCMIDADSRPKFELAAEFSRWARDPQRY 832
Qy 987 VITQNEB-LGPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPAGAGGVVH 1045
Db 833 LVIQGDRMKLPSPNDKSFQNLLEDEEDLMDABEYLV-QAENIPPIYTSRTRIDS 891
Qy 1046 RHRSSSTRSGGDLTLGLPSESEAPRS--PLAP-SEGAGSVDFDGLGMGAAGLQSLP 1102
Db 892 NRNQFYVRDGGYAAEQGV-PMPYRAPGCIIPAPVAGGATAEITFCNGLTKQVATL 950
Qy 1103 THDPSPLQRYSEDPYPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREG 1155
Db 951 AKEDSSTORYSADPTVFIPERVIRGELDEDCYTPMRDKPTDYLPNVEENPVSRRKNG 1010
Qy 1156 PLPAA-RPAGATLERAKTISPGKNVGVKDF-----AFGAVENPEYLTPOGGAAPQ 1206
Db 1011 DLQAVDNPVHN-----APNQPKAEDEYVNEPLYLNTFANTLENAEYL-----K 1055
Qy 1207 PHPPPAFSAFONLYWDQDPPPERGA--PPSTFKGTPT-----AENPEYL 1249
```

```
Db 1056>NNLPEKAKAFDNDPYMNHSLPPRSTLQHPDYLOEYSTKYFYKQNGRIRDVAENPEYL 1114
RESULT 6
P79754
ID AC P79754 PRELIMINARY; PRT: 1328 AA.
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE EzbB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=99177347; PubMed=10077531;
RX Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258 (1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
Query Match 34.0%; Score 2315; DB 13; Length 1328;
Best Local Similarity 40.6%; Pred. No. 1.2e-165;
Matches 522; Conservative 155; Mismatches 419; Indels 190; Gaps 31;
Qy 9 WGLLALLPP--GAASTQ---VCTCTDMKRLRASPETHLDMRLHLYQCGVQGNLEL 62
Db 4 WRILMCVASRLRAASSQTOEAVCPGTQNGLSSTGSOENQYNLKNORYKGEIIMGNLEI 63
Qy 63 TYLPTTNASLSFLQDIOEVQYVLIHQNVRQVPLQRLRIVRGTLQDPEDNYALAVLDNGDP 122
Db 64 TQIESNWDSEFLKTRITREVTGYVLIANNHFQEIPLGQLRVIRGNSLYERRFALSFLN--- 120
Qy 123 LNNQYTKANSKFTGITELQLRSLTEILKGGVLITQPNQLCYQDTILKWKDIFHKNNQAL 182
Db 121 ----YPKDGS--GLNQLGLMNLTEILDGVQIINNLYRYPGWVYWRDII-RNNDAPIE 173
Qy 183 LIDTNRSRACHPCSPMKCSRCSGESSEDCQSILTRTVACGCC-ARCKGPLPTCCHEQCA 241
Db 174 IOPNGERGVC---KSC-GNYCMGPKGQDQQLTKTVCAQCNCNDRCFGTSPRCCCHIECA 229
Qy 242 AGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPY 301
Db 230 AGCKGPLDTCFACRLFNDSGACVPCQPTLIYNKQTFQWETPNNAKYQVGSICVSOCP 289
Qy 302 NYLSTDVSGCTLVCPHLNQEVTAEDGTQ-CEKSKPCARVCYVGLGMEHLREVRVAVTSAN 360
Db 290 HFV-VDGSVCVSPDPKMEV--ERGSQRCCLCSGLCPKVCBGTGAE---QRQTVDSN 343
Qy 361 IQEFACKKIFGSLAPLPSFDGDPASNTAPLOPEOLQVPELEBITGYLYISAWPDSL 420
Db 344 IDSFNCTKIQSLHFLVTGILGDDPKNVPPLDAKKLEVFRVREITDILNIQSWPKELN 403
```



```
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLSRELGLALIIHHTHLCFVHTV 480
Db 404 DLSVFSSLTIOGRSLFKRFSLMVMRIPTLTSLGLSLREISDGSVYISQNAHLCHVHTV 463
QY 481 PWDQLFRNPH-QALLHTANRPEDECVEGLGLACHOLCARGHCWGPPTQCVCNCSQFLRGQE 539
Db 464 NWTQLFRGSRVRANSLSNRPAECVADGRVCDPLCSDSGCWGPGDQCLSCRNYSRHT 523
QY 540 CVEECRVLOGLPREVYNARH-CLPCHPEQOPQNGSVTCFPGPADQCVAHAHYKDPFPCA 598
Db 524 CVAGCHFNSGIPREFAGLNGVCVACHPECKPOTGKASCTGPGADECMACTKFRDGPYMS 583
QY 599 RCPGSGVKPDLSPYMPIKFPDEGACOPCPINCTHSCVDLDDGCPAERASPLTSISVAV 658
Db 584 SCPAGVN-DGEKGLIFKPNRRECHCEPCQNCTQGCGPGLNDC-----LEAARLTSSGOI 639
QY 659 VGILLVVLVGVVF-----GILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQ 711
Db 640 TGIALGVPAGLIFCLVFLGLMYHRLAIRKRANRRYLGESGEFELGP-GEKGTKVH 698
QY 712 MRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 771
Db 699 ARILKPSDLRKIKPLGSGVGTGSKGFWIPEGETVKIPVAIKTIQDSSGRQTFTEITDHL 758
QY 772 YVMAGVGSYPVSRLLGICLTSTVOLVTQLMPYCCLLDHYENRGRGLSGODLLNWCQIAK 831
Db 759 LSMGLDHPHYVRLGICPGTCLQVTLQSSHGSLEHROHKTSLDPORLNWCQVQIAK 818
QY 832 GMSYLEDVRLHRLDAARNVLKSPNHVKITDFGLARLLDIDETEVHADGCKVPKIKMAL 891
Db 819 GWYILEHRVHVHKNLAARNILKNDYQVOISDYGVADLLYPDDKVVYSETKTPKIKMAL 878
QY 892 ESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDV 951
Db 879 ESILFRYTHQSDVMSYGVTVWEMMSFGAEPYASVQPEVPSVLEKGERLSQPAICTIDV 938
QY 952 YMIWVKMIDSECPREFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLED 1011
Db 939 YMVWVKMIDENIRPTFKELASDFTRMARDPRYLIVIRMEG-----BD 982
QY 1012 DMGDVDAEYLVPOQGFCCPDAPAGGVMVHRHRSSTRSGGDLTLGLEPSEEEAP 1071
Db 983 SCMGFEFL-----RCGER---GLLEADLEDEEE--- 1008
QY 1072 RSLAPSEGAGSDVFDGLMG----AAKGLQLPHTDPSPLQ-----RYSEDPTV 1118
Db 1009 -----GLDRFATPSLQSPSPSWSTSPQINSVMVMTQLRYD----- 1044
QY 1119 PLPSETDGVAPLTCSP-EPVNO-----PDVRRPQPPSPREGPL--PAAR 1161
Db 1045 -FAVSQGGHIGYLPMSPPVDITRQMWYORSLSVRTLPRSAFRRSREAELEDGDAQ 1103
QY 1162 PAGATLERAKTLSPGKNGVVDVFAFGGAVENPEYLTPOGGAAPQHPPPAPSPAFDNL 1221
Db 1104 CAGIFRVR-----FGSERGN-----POGG----- 1122
QY 1222 YWDQPPPERGAPPSTFKGTPTAENPE 1247
Db 1123 --QQRKLSTASSPSPSKTWADEDE 1146
```

RESULT 7

```
Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGRF.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
```

```
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D8967724F07 CRC64;
```

```
Query Match 30.1%; Score 2047.5; DB 5; Length 1433;
Best Local Similarity 32.7%; Pred.No.2.2e-145;
Matches 473; Conservative 196; Mismatches 393; Indels 385; Gaps 36;

QY 26 CTGTDMLKRLPASPETHLDMLRHLYGCGCQVQVQCNLELTYPNTNASLSFLQDIOEVQGYVL 85
Db 1 CIGTNGRMSVPANREHYKYLDRYTNCTYVDGNLEITWIONITDNLFLQIHREVTGYVL 60
QY 86 IAHNQVRQVPLORLRIVRGTQLF-----EDNTALAVLDNGDPLNNQYIKANSKPIGTEL 140
Db 61 ISLYLDLPQVILPRQLIIRGRTTFKLNKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTEILKGVLIQRLQPCVQDTILWKDI-FHKNQQLALTLIDNRSRACHPCSPMC 199
Db 105 ELPALRDLGGSVGFNFNNLCHMKSIINWEEILLAPQTSMQYTFNFSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLTRTVACAGCA--RCKGPLPTDCCHCCCAAGCTGPKHSDCLACLH 257
Db 165 EVG-CWGEAHNCQRFSLKNCSPQSCGRFCGPKPRECCHLFCAGGCTGTQSDCLACKN 223
QY 258 FHSIGICELHCPALVTYNTDTFESHPNPEGRYTFGASCVTACPYNYLSTDVSGCTLVCPPL 317
Db 224 FYDDGVCKQCECPPMQIYNPTNYFWEPNPDGKYATCVRKCP-EHLLKDNAGACVRKCPK 282
QY 318 HNQEVTAEQGTQCEKCKPKCARVCYGLGMEHLREVRVTSANIQEFACKKIFGSLAEL 377
Db 283 GKMPQNSE-----CVPCGVCPKTCPEGIVH-----SNIGNYKDKCTIEGSLLEIL 339
QY 378 PESFDGDPASNT-----APLOPEQLQVFTLEBEITGYLYISAWPDSLPLDSVFQNLQ 429
Db 330 DQSPDFGQQVYTNFSFGPRYIKIDPDRLEVFSTVKEITGFINQAHHPNFTLTLYNFRNLE 389
QY 430 VIRGILHNGAY-SLTLOGLGISWGLRLSRELGLALIIHHTHLCFVHTVVPDQFLRN 488
Db 390 VVGGRLKENLFASVYIVKTSLSKLSLKLKSVNSGSIILENSDLCFVEDIDWSIEKKS 449
QY 489 PHQALLHTANRPEDECVEGLACHOLCARGHCWGPPTQCVCNCSQFLRGQECVEECRVLQ 548
Db 450 SDHEVWVQKNRNATECHEEGMECSECSKAGCKGKGPQCECKNKKYKGLDCK--- 506
```



```
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.7e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 60

QY 949 IDVYIMVWCWIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008
Db 61 IDVYIMVWCWIDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDAEEYLVPOOGFCPPAPGACGMVHRRSSSTRSGGDLTLGLEPSEE 1068
Db 121 LEDDDMGDLVDAEEYLVPOOGFCPPALGTGSTAHHRRSSARSGGELTLGLEPSEE 180

QY 1069 EAPRSPAPSEAGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDTPVLPSETDGYV 1128
Db 181 EAPRSPAPSEAGSDVFDGDLAVGVTKGLQSLSPHDLSPLOQRYSEDTPVLPSETDGYV 240

QY 1129 APLTSPQPEYVNPQDVRQPSREGPLPAARPAGATLERAKTLPSPKNGVVKVDFAFG 1188
Db 241 APLTSPQPEYVNPQDVRQPSREGPLPAARPAGATLERAKTLPSPKNGVVKVDFAFG 300

QY 1189 GAVENPEYLTPOGGAPOPHPAFSPAFDNLNLYNDODPPEPCAPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLTPOGGAPOPHPAFSPAFDNLNLYNDODPPEPCAPSTFKGTPTAENPEY 360

QY 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAGS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFEID63 CRC64;

Query Match      25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.3e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

QY 569 PONGSVTCFGEADQCVACAHYKDPFCFVARCPGKVPKPSLVNPIWKFPEEGACQPCPI 628
Db 141 PBEATPTKTPG--DHCMKCAHFDGPHCVKACPAVGLGENDTL-VNKYADANAVCQLCHP 197

QY 629 NCTHSCVDLDDKGCAPAEQASPLTSTVSAVV-GILLAVVLGVVFGTILIKRQOKIRKVTM 687
Db 198 NCTRCKGGLGEGCP---NGSKTPSIAAGVVGGLCLVVGIGLYLRRR-HIVKRTL 253

QY 688 RRLLOETELVELPTSPGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVK 747
Db 254 RRLQERELVELPTSPGEAPNOAHLRLAKETEFKKVKVLGSGAGFTVYKGLWIPGEKVK 313

QY 748 IPVAIKVLRNTSPKANKIILDEAYVMAGVGSPPYVSRLLGICLCTSTVQLVQLMPYGCCL 807
Db 314 IPVAIKELREATSPKANKIILDEAYVMASVDNPRVCRLLGICLCTSTVQLITQLMPYGCCL 373

QY 808 DHVRENRRGLSGODLLNWCQIAKMSYLEDLVRLVHRDLAARNVLKSPNHVKITDPGLA 867
Db 374 DVIHKONIGSQYLLNWCQIAKGNVLEERLVRDLAARNVLKTPHVKITDPGLA 433

QY 868 RLDDIDETEHADGGKVPKMMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI 927
Db 434 KILGADERKEYHAEGKVPKMMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI 493

QY 928 AREIPDLLEKGERLPQPPICTIDVYIMVWCWIDSECRPRFRELSEFMRMARDPQRFV 987
Db 494 ASEISVLEKGERLPQPPICTIDVYIMVWCWIDADSRPKRELIAEFKMDARDPPRYL 553

QY 988 VIQ-NEDLGPASPLDSTFYRSLLEDGDLVDAEEYLVPOOGFCPPAPGAGGMVHHR 1046
Db 554 VIQGERMHLPSPTDSKFYRTLMEBEDMEDIVDAEYLVPHOGFF----- 598

QY 1047 HRSSSTRSGGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQGH 631

QY 1102 PTHDPSPLQRYSEDTPVLPSET--DGYVAPITCSPQPEYVNPQDVRQPSREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGNFLESSIDGFL-----PAPEYVQ--LMPKKPS----- 675

QY 1160 ARPAGATLERAKTLPSPKNGVVKDFP-----AFGGAVENPEYL 1197
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Tyrosine-protein kinase.
DR NON TER
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF25E1 CRC64;

Query Match 25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.2e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

Qy 578 GPEADQCVACAHYKDPFCVCARCPGVKPDLSYMIWKFDEBACQPCPINCTHSCVDL 637
Db 1 GP--DHCKCAHFIDGPHCVKACGAGVLGENDTL-VMKYADANAVCQCHNCPTCKGKGP 57
Qy 638 DDKGCPAEQASPLTSIVSAV-VGILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSTAAGVVGGLCLLVVGLGIGLYLRRR-HIVKRTLRLRLQEREL 113
Qy 697 VEPLTPSGAMPNOAQMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTPSGEAPNOAHLKETEFKVKVLGSGAFGVYKGLWIPEGEKVKIPVAIKELR 173
Qy 757 ENTSPKANKEILDEAYVMAGVGSYPVSRLLGICLTSTVOLVQLMPYGCILLDHVRENRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQITQLMPYGCILLDIYREHKN 233
Qy 817 LGSQDLNWCQIAKMSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLIDIDE 876
Db 234 IGSQDLNWCQIAKMSYLEDVRLVHRDLAARNLVKTPQHVKITDFGLAKLLGADEKE 293
Qy 877 YHADGCKVPIKWALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 294 YHAEGGKVPKWALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISVLE 353
Qy 937 KGERLPPOPICTIDVYIMVWKWIMIDSECRPRFRELVSFESRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPPOPICTIDVYIMVWKWIMIDADSRPKRELIAEFSKWARDPPRYLVVLCQDERMH 413
Qy 996 PASPLDSTFYRSLLDDMDGLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSG 1055
Db 414 LPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTLCLEPSEAEAPRSL-----APSEGAGSDVFDGLMGAAKGLQSLPTHDPSPQLQ 1110
Db 455 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSEFVQ 491
Qy 1111 RYSEDPTVPUPSET--DGYVAPLTCSPOPEYVNPQVRPPOPPSPREGPLPAARAGATLE 1168
Db 492 RYSDPTGNFLRESIDGFL-----PAPEYVNO--LMPKKPS----- 526
Qy 1169 RAKTSPGKGVKVDVF-----AFPGAVENPEYL 1197
Db 527 ----TAMVQIQYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 12
Q8WYVO
ID Q8WYVO PRELIMINARY; PRT; 412 AA.
AC Q8WYVO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 9.2e-120;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

Qy 889 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60
Qy 949 IDVYIMVWKWIMIDSECRPRFRELVSFESRMARDPQRFVVIQNEDELGSPASPLDSTFYRSL 1008
Db 61 IDVYIMVWKWIMIDSECRPRFRELVSFESRMARDPQRFVVIQNEDELGSPASPLDSTFYRSL 120
Qy 1009 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLCLEPSEE 1068
Db 121 LEDDDMGDLVDAEYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLCLEPSEE 180
Qy 1069 EAPRSPALPSEGAGSDVFDGLMGAAKGLQSLPTHDPSPQRYSEDPVPLPSETDGYV 1128
Db 181 EAPRSPALPSEGAGSDVFDGLMGAAKGLQSLPTHDPSPQRYSEDPVPLPSETDGYV 240
Qy 1129 APLTCSQPPEYVNPQVRPPOPPSPREGPLPAARAGATLERAKTSLPGKGVKVDVFAFG 1188
Db 241 APLTCSQPPEYVNPQVRPPOPPSPREGPLPAARAGATLERAKTSLPGKGVKVDVFAFG 300
Qy 1189 GAVENPEYLTPOGGAPOP-----HPPPA---FSPATFDNL 1220
Db 301 GAVENPEYLTPOGGAALSPDLLPSAQPSSTITGTRTHQSGGLHPAPSKGHLRQRTQST 360
Qy 1221 YVWD-QDPPER-----GAPPSTFKGTPTAEN 1245
Db 361 WWTCCOPEPGQVRRSPDVSSSGSREGLTSAGIKRWEGPPTISRGTCCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE GAG, v-erb-A, v-erb-B protein.
GN GAG, v-ERB-A, v-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Buskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
```

RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk pkinaase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Strhmn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; ZF-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; ZNF C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 6.9e-116;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHPEQC 568
Db 354 IEKQESYLLAFEHYINVRKNIHPFWSKLLMKVADLRMIGAYHRSFLHMKVCEPTELS 413
QY 569 PONGSVTCFGEADQVACAHYKDPFCVACPSGVKPDLSYMPWKPDDEGACQPCPI 628
Db 414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHP 465
QY 629 NCTHSCVDLDDKGPAAQASPLTSIVSAVV-GILLVVVLGVVFGVILIKRQOKIRKYTM 687
Db 466 NCTRCCKGPGLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLYLRRR-HIVRKTL 521
QY 688 RRLQETELVPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVK 747
Db 522 RRLQERELVPLTPSGEAPNOAHLRLKETEFKVKVVLGFGAGFTVYKGLWIPEGEKVT 581
QY 748 IPVAIKVIRENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVOLVTLQMPYGCIL 807
Db 582 IPVAIKUREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMYPGCLL 641
QY 808 DHVRENRRGLSGQDLLNMCQIAGMSYLEDLVRLVHRDLAARNVLVKSPPNHVKITDFGLA 867
Db 642 DYIREHKNIGSQYLLANVCQIAKGMNLYEERHVRDLAARNVLVKTPOHVKITDFGLA 701
QY 868 RLLDIDETEHADGKVPDKMALESILRRRTHQSDVMSYGVTVVWELMTGAKPYDGIP 927
Db 702 KQLGADEKEYHAEGKVPDKMALESILHRIYTHQSDVMSYGVTVVWELMTGSKPYDGIP 761
QY 928 AREIPDLLEKGERLPQPPICITDVYIMVKCMIDSECRFRRELVSFESRMARDPQRFV 987
Db 762 ASEISSVLEKGERLPQPPICITDVYIMVKCMGSDSRPKFRELIAEFSEKMDRPPRYL 821
QY 988 VIQ-NEDLGASPLDSTFYRSLLEDVDAEYLVPOQGFPCPDPAFCGAGGVHHR 1046
Db 822 VIQGERMHLSPSTDSKYRFLMEEDMEDIVDAEYLVPHQGF----- 866
QY 1047 HRSSTSTRGGGDLTLGLPEPSEERAPRSLAPSEAGSDVFDCLGMGAAGLQSLPTHDP 1106
Db 867 -NSPST-----SRTPLLSSLSATS-----NSATKCIDRNGGH-- 898

QY 1107 SPLQRYSEDPVTPLPSETDGYVAPLTCSQPQYVNVQDVRPQPPSPREGPLPAARPAGAT 1166
Db 899 -----EVREDGFL-----PAPEYVNVQ--LMPKKPSTAMVQNOIYNYISLT 936
QY 1167 -LERAKTILSPGKNGVVKVDVFAFGGAVENPEYL 1197
Db 937 AISKLPMSRYQN-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scortling P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; 1FCG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.3e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPFCVACPSGVKPDLSYMPWKPDDEGACQPCINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPCTRCCKGP 57
QY 638 DDKGPAAQASPLTSIVSAVV-GILLVVVLGVVFGVILIKRQOKIRKYTMRLLOETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVGIGLYLRRR-HIVRKTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNOAHLRLKETEFKVKVVLGFGAGFTVYKGLWIPEGEKVTIPVAIKEL 173
QY 757 ENTSPKANKEILDEAYVMAGVGSPPVSRLLGICLTSTVOLVTLQMPYGCILDHVRENRR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVOLITQLMYPGCLLDYIREHKN 233
QY 817 LGSQDLLNMCQIAGMSYLEDLVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNMCVQIAKGMNLYEERHVRDLAARNVLVKTPODKITDFGLAKQLGADEKE 293
QY 877 YHAGGKVPDKMALESILRRRTHQSDVMSYGVTVVWELMTGAKPYDGIPAREIPDLLE 936
Db 294 YHAGGKVPDKMALESILHRIYTHQSDVMSYGVTVVWELMTGSKPYDGIPASEISSVLE 353
QY 937 KGERLPQPPICITDVYIMVKCMIDSECRFRRELVSFESRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICITDVYIMVKCMGSDSRPKFRELIAEFSEKMDRPPRYLVIOGDERMH 413

```

QY 996 PASPLDSTFYRSLLDDDDGLVDAAEYLVPOQGFCDPAPAGAGMVHRRSSSTRSG 1055
DB 414 LPSPTDSKEYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST--- 454
QY 1056 GGLTLGLLEPSEEAARSP-----APSEGAGSDVFDGLGMAAGLQSLPHDPSPLQ 1110
DB 455 -----SRTFLLSLSLTSATNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDCGVAPLTCSPOEYVQDVRPQPPSPREGPLPAARPAAGAT-LER 1169
DB 482 -----PVREDGFL-----PAPEYVQ--LMPKPSMTAMVOIQYINYSLSAISK 523
QY 1170 AKTSPGKNGVNDVFAFGAVENPEYL 1197
DB 524 LPMDSRYQN-----SHSTAVDNPEYL 544

```

RESULT 15

```

Q9WVF5
ID Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVITAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoro K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleichmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakrali L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BA023688.1; -.
DR EMBL; AK004883; BA023641.1; -.
DR EMBL; AK004911; BA023662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

```

```

Query Match 22.3%; Score 1519.5; DB 11; Length 655;
Best Local Similarity 44.0%; Pred. No. 5.1e-106;
Matches 281; Conservative 100; Mismatch 235; Indels 23; Gaps 7;

QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASDPETHLDMLRLHLYGCGVQGNLETLTYLPTN 68
DB 14 LLTALCAAGGALEKKVCGQTSNRLTQLGTFEDHFLSLQRMVNNCEVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIQEVQGVYLIHNOVROVPLQRLRIVRGTOLPEDNYALVLDNGDPLNNQYI 128
DB 74 YDLSFLTKIQEVAGYVLIATNTVERIPLENLQIIRGNALYENTYALAILSN----- 124
QY 129 KANSKFIGITELQLRSLTEILKGGVLIQRNPOLCYQDTILWKDI-----FKNNQLALTLI 184
DB 125 -YGTNRTGLRELPMRLQELILGAVRFSNNPILCNMDTIOWRDIQVNFMSNMSMDL--- 180
QY 185 DTRNSRACHPCSPMKGSCWGESSEDCQSLRTRTVCAAGCA-RCKGPLPTDCHEOCAAG 243
DB 181 -QSHFSSCPKCDPSCPNCGWGGEENCKLTKIIQAQCSHRCSRSPSCCHNCAAG 239
QY 244 CTGPKHSDCLACLHFNHSGICELHCPALVYNTDTPESMPNPGRYTFGASCVTACPNY 303
DB 240 CTGPRESCLVCQKFODEATCKDTCPPLMLNPTTYQMDVNPBGKYSFGATCVKKCPRY 299
QY 304 LSTDVGSCTLVCLHNEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSANIOE 363
DB 300 VVTDHGSVCVRACGPDYEV-EEDGIRKCKKCDGCPCKVCNGIGIGEFKOTLSINATNIKH 358
QY 364 FAGCKKIFGSLAPESPFGDPSASNTAPLOPQLOVFETLEETLYIYISAWPDSLPDLS 423
DB 359 FYCTAISGDLHLPLVAFKGSFTRTPPLDPRLEILKTVKEITGELLIQAWPDNWDLH 418
QY 424 VFQNLQVIRGRILHNGAYSLTLQGLISWGLRSLRELGLALIHNTLHLCFVHTVPWD 483
DB 419 AFENLEIIRGRTKHQGFSLAVVGLNITSLGRSLKEISDGDVILSGNRNLVCANTINWK 478
QY 484 QLFRRPHQALLHTANRPEDECVGEGELACQLCARHCWGPPTQCVCNCSOFLRGQECVEE 543
DB 479 KLFGTPNQTKIMNRAEKDKAVNHVNCPLSCSECGWGPEDCVSCQNVSRGRCVEK 538
QY 544 CRVLOGLPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPDFVCARCPSG 603
DB 539 CNILEGEPREFVENSECICHPCECLPQAMNITCTGRGPDNCIOCAHYIDGPHCVKTCPCAG 598
QY 604 VKPDLSSYPMIWKPPDEEGACQPCINCTHSCVDLDDKGC 642
DB 599 IMGENNTL-VWKYADANNVCHLCHANCTTYGCAGPLQGC 636

```

Search completed: July 22, 2003, 09:00:17
Job time : 54.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds
(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-103-117-12
Perfect score: 6809
Sequence: 1 MELAAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A. Geneseq, 101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6736	98.9	1255	21	AA1992620 Human heregulin 2
2	6736	98.9	1255	22	AAE12130 Human tyrosine kin
3	6736	98.9	1255	22	AAB60167 HER2 transgene pla
4	6736	98.9	1255	23	AAU74545 Human HER2 (ErbB2)
5	6730	98.8	1255	17	AAW01111 HER-2/neu protein.
6	6730	98.8	1255	20	AAW92406 Human HER-2/neu on
7	6730	98.8	1255	21	AAB21198 Human HER-2/neu pr
8	6730	98.8	1255	21	AA1984780 Amino acid sequenc
9	6730	98.8	1255	22	AAB85458 Human HER-2/neu pr
10	6730	98.8	1255	22	AAG88267 HER2/neu amino aci

11	6730	98.8	1255	23	AAE24067 Human Her-2 protei
12	6730	98.8	1255	23	AAE20479 Human Her-2/neu pr
13	6730	98.8	1255	23	AAAM51143 Human Her-2/neu on
14	6730	98.8	1255	23	AAU77114 Human Her-2/neu po
15	6687	98.2	1433	14	AAR39568 Sequence of c-erbB
16	6566	96.4	1223	23	AAU98923 Human breast cance
17	6413	94.2	1200	21	AAB21208 Human HER-2/neu pr
18	5967.5	87.6	1256	21	AAB21199 Rat Her-2/neu onco
19	5967.5	87.6	1256	23	AAAM51144 Mouse Her-2/neu pr
20	5939.5	87.2	1256	21	AAB21206 Mouse Her-2/neu pr
21	5939.5	87.2	1256	22	AAG62860 Amino acid sequenc
22	5939.5	87.2	1256	23	AAAM51151 Mouse Her-2/neu on
23	4816	70.7	919	21	AAB21203 Human HER-2/neu fu
24	4089.5	60.1	920	23	AAAM51148 Her-2/neu extracel
25	4089.5	60.1	926	23	AAAM51153 Mouse Her-2/neu ex
26	3700	54.3	712	21	AAB21204 Human HER-2/neu fu
27	3700	54.3	712	23	AAAM51149 Her-2/neu extracel
28	3554	52.2	782	18	AAW19764 Her2-GM-CSF immuno
29	3552	52.2	653	21	AAB21200 Extracellular HER-
30	3552	52.2	653	23	AAAM51145 Human ErbB2 oncopr
31	3514	51.6	645	22	AAB60408 Human ErbB2 extrac
32	3514	51.6	645	22	AAB61593 DC9ScFv-erbB2EC fu
33	3449	50.7	951	21	AA144993 Extracellular port
34	3349	49.2	624	11	AAAR08222 Amino acid sequenc
35	3160	46.4	1210	21	AAB19259 Human EGF receptor
36	3160	46.4	1210	21	AAAY50616 Human Her-1 protei
37	3160	46.4	1210	23	AAE23019 Human epidermal gr
38	3160	46.4	1210	23	AAAM50768 Amino acid sequenc
39	3160	46.4	1210	22	AAB68420 Human epidermal gr
40	3158	45.8	1210	23	ABP51768 Human protein for
41	3119	45.3	583	23	AAE20483 Human protein for
42	3084	45.3	587	23	AAE20481 Human protein for
43	3084	45.3	589	23	AAE20484 Human protein for
44	3083	45.3	600	23	AAE20482 Human protein for
45	3083	45.3	600	23	AAE20482

ALIGNMENTS

RESULT 1
AA1992620
ID AA1992620 standard; Protein; 1255 AA.
XX AC AA1992620;
XX AC AA1992620 (first entry)
XX DT 10-AUG-2000 (first entry)
XX DE Human heregulin 2 (Her2).
XX KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX OS Homo sapiens.
XX FH Key
FH Domain Location/Qualifiers
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"

FT Region /label= Cysteine_rich_domain
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 230..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= C-terminal_domain
 FT XX
 PN WO200020027-A2.
 XX 13-APR-2000.
 PD 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 PR 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI: 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 PS This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T-helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6736; DB 21; Length 1255;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
 QY 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLRPASPETHLDMRLHLYQGCQVVGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLRPASPETHLDMRLHLYQGCQVVGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVGTQLFEDNVALAVLDNG 120
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVGTQLFEDNVALAVLDNG 120
 QY 121 DPLNNQYIKANSKFITELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
 DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMKSGRCSESDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPMKSGRCSESDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
 QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNEGRTTFCASCVTACP 300
 DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPNEGRTTFCASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 QY 361 IOEFAGCKKIFGSLAFLPESDGDPAANTAPLOPEQLQVFETLEITGYLISANPDSL 420
 DB 361 IOEFAGCKKIFGSLAFLPESDGDPAANTAPLOPEQLQVFETLEITGYLISANPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELCSGLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELCSGLALIHNNTHLCFVHTV 480
 QY 481 PWDQLFRNPQALLHTANRPEDECYVGEGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQEC 540
 DB 481 PWDQLFRNPQALLHTANRPEDECYVGEGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQEC 540
 QY 541 VEECKVLOGLPREYVNAHCLPCHPECCOPQNSVTCFGEADOCVACAHYKDPFCVARC 600
 DB 541 VEECKVLOGLPREYVNAHCLPCHPECCOPQNSVTCFGEADOCVACAHYKDPFCVARC 600
 QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQORASPLTSIVSAVVG 660
 DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQORASPLTSIVSAVVG 660
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIKTEL 720
 DB 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIKTEL 720
 QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEATVMAGVGP 780
 DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEATVMAGVGP 780

```
QY 781 YVSRLLGICLTSTVOLTPYQCLLDHVENRGRGLGSQDLLNWCMIAGKHSYLEDRV 840
Db 781 YVSRLLGICLTSTVOLTPYQCLLDHVENRGRGLGSQDLLNWCMIAGKHSYLEDRV 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKVMWMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKVMWMALESILRRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDECEPRPRELVSEFSRWARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDECEPRPRELVSEFSRWARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVLPQOGFCFDPAPGAGMVHRRHSSTSSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVLPQOGFCFDPAPGAGMVHRRHSSTSSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 774..782
FT /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
DR N-PSDB; AAD19731.
XX
XX Novel synthetic therapeutic compound for inducing immune response and
PT for use in adoptive immunotherapy, has enhanced binding to major
PT histocompatibility molecules and enhanced immunoregulatory properties
PT .
```

```
XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
Query Match 98.9%; Score 6736; DB 22; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
QY 1 MELAALCRMGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYGQCQVVOGNL 60
Db 1 MELAALCRMGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLYGQCQVVOGNL 60
QY 61 ELYLPTNASLSFLQDIOEVQGVYLAHNQVROVPLORLAIVRGTQLFENYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIOEVQGVYLAHNQVROVPLORLAIVRGTQLFENYALAVLDNG 120
QY 121 DPLNNQYIKANSKFIGITELQLRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHNKNOLA 180
Db 121 DPLNNQYIKANSKFIGITELQLRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHNKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPFHQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPFHQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLVFETLEEITGYLYISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLVFETLEEITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGSLRELGSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSILTLQGLGISWLGSLRELGSGLALIHNTLHLCFVHTV 480
QY 481 PWDOLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRQEC 540
Db 481 PWDOLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVCNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECQPNQSGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQPNQSGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPDEBEGACQPCINCTHSCVDLDDKGCPEAQASPLTSIVSAVG 660
Db 601 PSGVKPDLSPYMPIWKFPDEBEGACQPCINCTHSCVDLDDKGCPEAQASPLTSIVSAVG 660
QY 661 ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
```


Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHYHADGKGVPIKWMALLESILRRRFT 900
Qy 901 HQSDVMSYCVTWELMTGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVYMINVKWM 960
Db 901 HQSDVMSYCVTWELMTGAKPYDGI PAREIPDLLEKGERLPDPPICTIDVYMINVKWM 960
Qy 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
Qy 1021 BEYLVPQGGFTCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEAEPRLAPSEG 1080
Db 1021 BEYLVPQGGFTCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEAEPRLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPQ 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDODPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 4

AAU74545
ID AAU74545 standard; Protein; 1255 AA.
AC AAU74545;
XX
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoelic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
XX 16-MAR-2001; 2001US-0811123.
XX
XX 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
XX (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
XX Erickson S, Schwall R, Sliwkowski M;
PI WPI; 2002-163686/21.
XX
DR N-PSDB; ABK14058.
XX
XX

Treating tumour characterised by overexpression of epidermal growth factor receptor, ErbB or cancer in mammal, comprises administering anti-ErbB antibody-maytansinoid conjugate to the mammal -

Example 3; Fig 7; 93pp; English.

The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoelic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 98.9%; Score 6736; DB 23; Length 1255;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1243; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHDMLRHLRYQSCVQVQNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHDMLRHLRYQSCVQVQNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVROPVLPQRLRIVRGTLQDFEDNYVALVDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVROPVLPQRLRIVRGTLQDFEDNYVALVDNG 120
Qy 121 DPLNNQVIKANSKFIGITELQRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHKNQOLA 180
Db 121 DPLNNTPTVTCASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLWKDIFHKNQOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGGCARKGPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGGCARKGPLPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEBITGYLIISAMPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEBITGYLIISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSRLRELGSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRPHOALLHTANRPEDECVGEGCLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
Db 481 PWDQLFRPHOALLHTANRPEDECVGEGCLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIUKETEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIUKETEL 720
Qy 721 RKVKVLGSGAFGTGYKGIWIPDGENVKI PVAIKVLRENTSPKANKEILDEAYVNAVGVSP 780
Db 721 RKVKVLGSGAFGTGYKGIWIPDGENVKI PVAIKVLRENTSPKANKEILDEAYVNAVGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHRNENRGRIGSQDLNLCWQIAKMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALSIILRRRT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALSIILRRRT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMWKCMW 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMWKCMW 960
Qy 961 IDSECRPRELVSEFSRMRDPQRFVJIONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRELVSEFSRMRDPQRFVJIONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 ESYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSG 1080
Db 1021 ESYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVVRPQPPSPREGPLPAARPAATLERAKTSLSPGKNGVVDVFAFGGAVENPEYLTQ 1200
Db 1141 NQPDVVRPQPPSPREGPLPAARPAATLERAKTSLSPGKNGVVDVFAFGGAVENPEYLTQ 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111
ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= Intracellular domain
FT /note= "claimed domain, useful for immunisation"
XX
PN WO9630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
PA (UNITW) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI: 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX SQ Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 17; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASCTGTCTDMKRLPASPETHLDMRLHLVYQGCQVVGNL 60
Db 1 MELAALCRWGLLLALLPPGAASCTGTCTDMKRLPASPETHLDMRLHLVYQGCQVVGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIHANVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQYVLIHANVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKPFITELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTIIDTNRSRACHPCSPMKSGRSGESSEDCQSLTRTVTCAGGCARCKGPLPTDCHEOC 240
Db 181 LTIIDTNRSRACHPCSPMKSGRSGESSEDCQSLTRTVTCAGGCARCKGPLPTDCHEOC 240
Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNLSLTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFLEETITGVLYISAWPDSLP 420
Db 361 IOEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFLEETITGVLYISAWPDSLP 420
Qy 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVPQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Db 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
Qy 601 PSQVVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSQVVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEIILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDCIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960
 DB 901 HQSDVWSYGVTVWELMTFGAKPYDCIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960
 QY 961 IDSECRPRFRELVSERFMRDPPQFVVIQNEIDLGPAASPLDSTFYRSLLEDGDLVDA 1020
 DB 961 IDSECRPRFRELVSERFMRDPPQFVVIQNEIDLGPAASPLDSTFYRSLLEDGDLVDA 1020
 QY 1021 EYLVPQGGFCPPAPCAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080
 DB 1021 EYLVPQGGFCPPAPCAGGMVHRHRSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVDFDGLGMAAGKGLQSLTHDPSPQLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 DB 1081 AGSDVDFDGLGMAAGKGLQSLTHDPSPQLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKXGVVVKDVFAGGAVENPEYILTPQ 1200
 DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKXGVVVKDVFAGGAVENPEYILTPQ 1200
 QY 1201 GGAAPQHPPPAPFADNLYYWDQDPPERCAPPSTFKGTPTAENPEYILGDLVDPV 1255
 DB 1201 GGAAPQHPPPAPFADNLYYWDQDPPERCAPPSTFKGTPTAENPEYILGDLVDPV 1255

RESULT 6

AAW92406

ID AAW92406 standard; Protein: 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Region 676..1255

XX /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.
 XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

PS This sequence represents the human HER-2/neu oncogene protein. A fragment
 XX

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 20; Length 1255;
 Beat Local Similarity 98.9%; Pred. No. 0;
 Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
 DB 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVVOGNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTLQDPEDNYALAVLNG 120
 DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTLQDPEDNYALAVLNG 120
 QY 121 DPLNQVIKANSKFIGITELQLRSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
 DB 121 DPLNNTTPTVTCASPGGURELQLRSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACGGCARCKGPLPTDCCHQEC 240
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACGGCARCKGPLPTDCCHQEC 240
 QY 241 AAGCTGPKHSDCLACHLHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHLHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLGHEHLREVRVTSAN 360
 DB 301 YNYLSTDVGSLVCPPLHNOBVTAEADGTORCEKSKPCARVCYGLGHEHLREVRVTSAN 360
 QY 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQVLPETLEEITGLYISAWPDSL 420
 DB 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQVLPETLEEITGLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLGLAIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLGLAIHNNTHLCFVHTV 480
 QY 481 PWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 DB 481 PWDLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
 QY 601 PSGVKPDLSPYMPWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660
 DB 601 PSGVKPDLSPYMPWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660
 QY 661 ILLVVLGVVFGILIKRROOKIRKTYMRLLQSTELVEPLTPSCAMPNQAMRILKETEL 720
 DB 661 ILLVVLGVVFGILIKRROOKIRKTYMRLLQSTELVEPLTPSCAMPNQAMRILKETEL 720
 QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 DB 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
 QY 781 YVSRLLIGICLTSTVQLVTQMPYGCGLLDHVRENRGLGSDLLNWCNQIAKMSYLEDRV 840
 DB 781 YVSRLLIGICLTSTVQLVTQMPYGCGLLDHVRENRGLGSDLLNWCNQIAKMSYLEDRV 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRFT 900

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Dy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMGDLVDA 1020
Dy 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMGDLVDA 1020
Qy 1021 EYLVPQOGFFCPDAPGAGGMVHRHRSSTRSGGDLTLGLPSEEAAPRSPAPSEG 1080
Dy 1021 EYLVPQOGFFCPDAPGAGGMVHRHRSSTRSGGDLTLGLPSEEAAPRSPAPSEG 1080
Qy 1081 AGSDVFDGDLQMGAAKGLQSLTPHDPSPLQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Dy 1081 AGSDVFDGDLQMGAAKGLQSLTPHDPSPLQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPOPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Dy 1141 NOPDVRPOPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Dy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein; 1255 AA.
AC AAB21198;
XX 12-JAN-2001 (first entry)
XX Human HER-2/neu protein.
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
XX breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX colon cancer.
XX Homo sapiens.
XX W0200044999-A1.
XX 03-AUG-2000.
XX 28-JAN-2000; 2000WO-US02164.
XX 29-JAN-1999; 99US-0117976.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM.
XX Cheever MA, Gheysen D;
XX WPI; 2000-505976/45.
XX N-PSDB; AAA89736.
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
XX useful for vaccinating against breast, ovarian, colon, lung and
XX prostate cancers -
XX Claim 52; Fig 7; 128pp; English.
XX The present sequence is the human HER-2/neu protein. It is a member of
XX the tyrosine kinase family of receptor-like glycoproteins and shows
XX homology to the epidermal growth factor receptor (EGFR). It probably
XX plays a part in cell growth and/or differentiation. The HER-2/neu
XX gene is an oncogene. An HER-2/neu fusion protein comprising a
XX HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
XX domain may be used to treat or prevent cancer by eliciting or
XX enhancing an immune response to the HER-2/neu protein. It may be used
XX to treat malignancies such as breast, ovarian, colon, lung and
XX prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 21; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRLPASPETHLDMRLHLYGQCQVQGNL 60
Dy 1 MELAALCRWGLLLALLPPGAASQVCTGTDKMLRLPASPETHLDMRLHLYGQCQVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVYLI AHNVQVQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
Dy 61 ELTYLPTNASLSFLQDIQEVQGVYLI AHNVQVQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Dy 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTCCHEQC 240
Dy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAGGCARCKGPLPTCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQDTCRCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Dy 301 YNYLSTDVGSCTLVCPLNHNOEVTAEQDTCRCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Qy 361 IQBFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
Dy 361 IQBFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
Dy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLRELGLALIHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Dy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEBCRVLQGLPREVYNARHCLPCHPECOPONGSVTCFGEADQCVACHYKDPFCVCARC 600
Dy 541 VEBCRVLQGLPREVYNARHCLPCHPECOPONGSVTCFGEADQCVACHYKDPFCVCARC 600
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660
Dy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRROQKIRKYTMRLRLQSTELVEPLTPSGAMPNQAMRILKETEL 720
Dy 661 ILLVVLGVVFGILIKRROQKIRKYTMRLRLQSTELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDSEAYMAGVGSP 780
Dy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDSEAYMAGVGSP 780
Qy 781 YVSRLGICLSTVQLVLTQMPYGCLLDHVRENHGRGLSODLLNWCQIAKMSYLEDVR 840
Dy 781 YVSRLGICLSTVQLVLTQMPYGCLLDHVRENHGRGLSODLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Dy 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMGDLVDA 1020

Db 961 IDSECRPRELVSEFSRMARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1021 EYLVPOQGFCCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFCCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSESDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSESDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRPQPSREGPLPAARPAGATLERAKTLSPCKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPSREGPLPAARPAGATLERAKTLSPCKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1201 GGAAPOHPHPPAFSPAFDNLYYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPOHPHPPAFSPAFDNLYYWDQPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8

AA584780
ID AA584780 standard; Protein; 1255 AA.
XX
AC AA584780;
XX
DT 08-AUG-2000 (first entry)
XX
DE Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW tumor cell proliferation; tissue degeneration; arthropathy;
KW bone resorption; inflammatory disease; degenerative disorder;
KW wound healing.
XX
OS Homo sapiens.
XX
PN WO200020579-A1.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-CA00912.
XX
PR 02-OCT-1998; 98US-0165192.
XX
PA (UYMC-) UNIV MCMASTER.
XX
PI Muller WJ, Siegel PM;
XX
PP WPI; 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
PT erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPLICE erbb-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbb-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;

Query Match	98.8%	Score 6730;	DB 21;	Length 1255;				
Best Local Similarity	98.9%	Pred. No. 0;						
Matches 1241;	Conservative	3;	Mismatches	11;	Indels	0;	Gaps	0;
QY	1	MELAAALCRGGLLALLPPGAASTQVCTGDMKLRLPASPTHLDMLRLHLYQGQVVOGNL	60					
Db	1	MELAAALCRGGLLALLPPGAASTQVCTGDMKLRLPASPTHLDMLRLHLYQGQVVOGNL	60					
QY	61	ELTYLPTNASLFLQDIQEVQGVLIHAHNOVRQVPLQRLIRVGTOLFEONYALAVLDNG	120					
Db	61	ELTYLPTNASLFLQDIQEVQGVLIHAHNOVRQVPLQRLIRVGTOLFEONYALAVLDNG	120					
QY	121	DPLNNQYIKANSKFIGITELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA	180					
Db	121	DPLNNTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA	180					
QY	181	LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGGCARCKGPLTDCCHQEC	240					
Db	181	LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCCAGGCARCKGPLTDCCHQEC	240					
QY	241	AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300					
Db	241	AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300					
QY	301	YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCKSPCARVCYGLGMEHLREVRAVTSAN	360					
Db	301	YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCKSPCARVCYGLGMEHLREVRAVTSAN	360					
QY	361	IQEFACCKIFGSLAFLPESFDGDPASNTAPLOPEOLQVPETLEEITGYLYISAWPDSLP	420					
Db	361	IQEFACCKIFGSLAFLPESFDGDPASNTAPLOPEOLQVPETLEEITGYLYISAWPDSLP	420					
QY	421	DLVSFONLQVIRGRILHNGAYSILTQGLGTSWGLRSLRELGSGLALIHHTHLCFVHTV	480					
Db	421	DLVSFONLQVIRGRILHNGAYSILTQGLGTSWGLRSLRELGSGLALIHHTHLCFVHTV	480					
QY	481	PWDQLFRNPQHALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQEC	540					
Db	481	PWDQLFRNPQHALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQEC	540					
QY	541	VBEERVVLQGLPREYVNHARCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600					
Db	541	VBEERVVLQGLPREYVNHARCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600					
QY	601	PSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPA2ORASPLTISVSAVG	660					
Db	601	PSGVKPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKCPA2ORASPLTISVSAVG	660					
QY	661	ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720					
Db	661	ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720					
QY	721	RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELDEAYVMAGVSP	780					
Db	721	RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELDEAYVMAGVSP	780					
QY	781	VYSRLGICLTSTVOLVTOLMPYGCILLDHVRENHRLGSGQDLLNWCQIAGKHSYLEDDR	840					
Db	781	VYSRLGICLTSTVOLVTOLMPYGCILLDHVRENHRLGSGQDLLNWCQIAGKHSYLEDDR	840					
QY	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKNWMALESILRRRFT	900					
Db	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKNWMALESILRRRFT	900					
QY	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYIMVKCMW	960					
Db	901	HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPICTIDVYIMVKCMW	960					
QY	961	IDSECRPRELVSEFSRMARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDMGLVDA	1020					
Db	961	IDSECRPRELVSEFSRMARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDMGLVDA	1020					
QY	1021	EYLVPOQGFCCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080					

Db 1021 EYLVPOQGFCCDPAPAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQLPHTDPSPLORYSEDPVLPSPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQLPHTDPSPLORYSEDPVLPSPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPPPAPSPAFDNLYYWDQPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAPSPAFDNLYYWDQPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 9
AA885458
ID AA885458 standard; Protein; 1255 AA.
XX AC AA885458;
XX 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
KW Homo sapiens.
XX WO200153463-A2.
PN 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
PF 21-JAN-2000; 2000US-0177545.
PR (CORI-) CORIXA CORP.
PA Cheever MA, Hand-Zimmermann S;
PI WPI; 2001-476112/51.
XX N-PSDB; AAH23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 98.9%; Score 6730; DB 22; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAAALCRWGLLLALLLPFGAASSTVCTGTDMKRLPASPEHLDMLRHLVQGCQVVGNNL 60
Db 1 MELAAALCRWGLLLALLLPFGAASSTVCTGTDMKRLPASPEHLDMLRHLVQGCQVVGNNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGVVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 1200

Db 61 ELTYLPTNASLSFLQDIEQVGVVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQRLSLTEILKGGVLIQORNPOLCYODTTLWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQLRLSLEILKGGVLIQORNPOLCYODTTLWKDI FHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLTDCCHQEC 240
Db 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLTDCCHQEC 240
Qy 241 AAGCTGPKGSDCLACLFHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLFHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGLSGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGLSGLALIHNTHLFCFVHTV 480
Qy 481 PWDOLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCWGPQTCVNCQFLRQEC 540
Db 481 PWDOLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCWGPQTCVNCQFLRQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFFCVARC 600
Qy 601 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEOBASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYPIWKFPDEEGACQPCINCTHSCVDLDDKCPAEOBASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKTEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKTEL 720
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCQIAKGMYSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLGSDLLNWCQIAKGMYSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEVHADGGKVPKMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEVHADGGKVPKMMALESILRRFT 900
Qy 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Db 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFRELVSFERSMARDPQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECRPRFRELVSFERSMARDPQRFVJONEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 EYLVPOQGFCCDPAPAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCCDPAPAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRPLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQLPHTDPSPLORYSEDPVLPSPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQLPHTDPSPLORYSEDPVLPSPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPO 1200

Db 1141 NQDVRPQPPSPREGPLPAARPAATLERPKTSLSPGKNGVVKVAFEGGAVENPEYJTPQ 1200
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
XX W0200141787-A1.
XX
XX 14-JUN-2001.
XX
XX 11-DEC-2000; 2000MO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
XX (EPIM-) EPIMMUNE INC.
PA
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
XX Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 22; Length 1255;

Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASSTOVCTGTDMKRLPAS PETHLDMLRHL YGCGVQVQNL 60
Db 1 MELAALCRWGLLALLPPGAASSTOVCTGTDMKRLPAS PETHLDMLRHL YGCGVQVQNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRIVRGTOFLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSDCSLTRTV CAGGCARCKGPLTCCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSDCSLTRTV CAGGCARCKGPLTCCCHEQC 240
Qy 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVTYNTDT FESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVTYNTDT FESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVTAE DGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLNHNOEVTAE DGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVP EETLEEITGYLYISAWPDSLP 420
Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQLOVP EETLEEITGYLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGIRLHNGAYSILTLOGLGISWLGSLR LRELGLSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGIRLHNGAYSILTLOGLGISWLGSLR LRELGLSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCAR GHCGWGPBPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCAR GHCGWGPBPTQCVNCSQFLRGQEC 540
Qy 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGE PADOCVACAHYKDPFCVARC 600
Db 541 VESCRVLQGLPREVYNARHCLPCHPECPONGSVTCFGE PADOCVACAHYKDPFCVARC 600
Qy 601 PSQVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDK GCPAEQASPLTISVAVVG 660
Db 601 PSQVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDK GCPAEQASPLTISVAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTP SGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTP SGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR ENTSPKANKEILDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR ENTSPKANKEILDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRGLS QDLDLNCWQIAKMSYLEVDV 840
Db 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRGLS QDLDLNCWQIAKMSYLEVDV 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG GKVPIKWMALLESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADG GKVPIKWMALLESILRRFT 900
Qy 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLEK GERLPPOPICTIDVYMIWVKWM 960
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLEK GERLPPOPICTIDVYMIWVKWM 960
Qy 961 IDSECPRELVSEFSRMDPQRFVVIQNE DLGPASPLDSTFY RSLLEDDDDMDGLVDA 1020
Db 961 IDSECPRELVSEFSRMDPQRFVVIQNE DLGPASPLDSTFY RSLLEDDDDMDGLVDA 1020
Qy 1021 EYLVFPQQGFCCDDPAPGAGGMVHRRHSSTSGGGDLTL GLSEPEEEAPRSPAPSE 1080
Db 1021 EYLVFPQQGFCCDDPAPGAGGMVHRRHSSTSGGGDLTL GLSEPEEEAPRSPAPSE 1080

```
Db 1021 EEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYLTQP 1200
Db 1141 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYLTQP 1200
Qy 1201 GGAAPQHPHPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein: 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowser LM;
XX
DR WPI; 2002-471192/50.
XX
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASSTVCTGCTDMKRLPASPEHLDMLRHLYGCCVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASSTVCTGCTDMKRLPASPEHLDMLRHLYGCCVVOGNL 60
```

```
Qy 61 ELTYLPTNASLFLQDIOEQVGYVLI AHNOVROVPLQRLRIVRGTTQLFEDNYALAVLNG 120
Db 61 ELTYLPTNASLFLQDIOEQVGYVLI AHNOVROVPLQRLRIVRGTTQLFEDNYALAVLNG 120
Qy 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWCSESDCQSLTRTVCAAGCARCKGPLPTCCCHQC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWCSESDCQSLTRTVCAAGCARCKGPLPTCCCHQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRATSAN 360
Db 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRATSAN 360
Qy 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPFETLEETGTYLISAMPDSL 420
Db 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLQEQLOVPFETLEETGTYLISAMPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGLRSLRELGLSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGLRSLRELGLSLALIHNTLHLCFVHTV 480
Qy 481 PDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARHGCMGPGPTQCVNCSQFLRGQEC 540
Db 481 PDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARHGCMGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Qy 601 PSGVKPDLSYMPKWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSYMPKWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVNAVGVSP 780
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDEAYVNAVGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLMPYGCCLLDHVRENRGLCSQDLNWCQIAKMSYLEOVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLMPYGCCLLDHVRENRGLCSQDLNWCQIAKMSYLEOVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRRFT 900
Qy 901 HQSDVMSYGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWKCM 960
Db 901 HQSDVMSYGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWKCM 960
Qy 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
Qy 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAKGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFAGGAVENPEYLTQP 1200
```

Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERRPKTUSPGKNGVVKDVAFPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSFPGKPTPTAENPEYLGDDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSFPGKPTPTAENPEYLGDDVPV 1255
RESULT 12
ID AAE20479
AC AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI: 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer
XX Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;
Query Match 98.8%; Score 6730; DB 23; Length 1255;

Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASQTQCTGTDMLKRLPASPEHLDMLRHLVQGCVOVGNL 60
Db 1 MELAALCRWGLLALLPPGAASQTQCTGTDMLKRLPASPEHLDMLRHLVQGCVOVGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNOVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNOVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLANNQYIKANSKFIGITELQLRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLANNTPVTGASPGGLRELQLRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNNOLA 180
QY 181 LTLIDTNRGRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRGRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCORCEKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCORCEKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420
Db 361 IOBFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGTSWGLSLRSLRELGLSLIHNHILCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGTSWGLSLRSLRELGLSLIHNHILCFVHTV 480
QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSQFLRQEC 540
QY 541 VEECRVLQGLPREYVNNARHCLPCHPECOPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNNARHCLPCHPECOPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEARASPLTSTVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEARASPLTSTVSAVVG 660
QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQSTELVEPLTPSGAMPNQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQSTELVEPLTPSGAMPNQMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDVDR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDVDR 840
QY 841 LVHRDLAARNVLKSGPNHVKIITDFGLARLLDIDETEHADGGKVPKIMWALESILRRRFT 900
Db 841 LVHRDLAARNVLKSGPNHVKIITDFGLARLLDIDETEHADGGKVPKIMWALESILRRRFT 900
QY 901 HOSDVMYSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIWKCMW 960
Db 901 HOSDVMYSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMIWKCMW 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
QY 1021 ESYLVPQOQFFCPDPAAGAGGWHHRHRSSTSRSGGDLTJLGLEPSEEEAPRSLAPSEG 1080

781 YVSRLLGICLTSTVQLVLTQMPYGCGLLDHVRENRGRGLSGDQLLNWCQIAKGNMYSLEDVR 840
|
781 YVSRLLGICLTSTVQLVLTQMPYGCGLLDHVRENRGRGLSGDQLLNWCQIAKGNMYSLEDVR 840
|
841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900
|
841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900
|
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 960
|
901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 960
|
961 IDSECRPRFRELVSERFMRDQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDEDDMGDLVDA 1020
|
961 IDSECRPRFRELVSERFMRDQRFVVIQNEIDLGSPASPLDSTFYRSLLEDDEDDMGDLVDA 1020
|
1021 EYLVPQOQFFCDDPAPGAGMWHHRSSSTSGGGDLTLGLERPESEEAAPRPLAPSEG 1080
|
1021 EYLVPQOQFFCDDPAPGAGMWHHRSSSTSGGGDLTLGLERPESEEAAPRPLAPSEG 1080
|
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
|
1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
|
1141 NQPDVREPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTQP 1200
|
1141 NQPDVREPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTQP 1200
|
1201 GGAAPOHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1255
|
1201 GGAAPOHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGDLVVPV 1255
|

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR WPI: 2002-280741/32.
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.8%; Score 6730; DB 23; Length 1255;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1241; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MELAALCRGGLLLALLPPGAASTQVCTGDMKURLPASPTHLDMLRHLYQGQVVOGNL 60
|
DB 1 MELAALCRGGLLLALLPPGAASTQVCTGDMKURLPASPTHLDMLRHLYQGQVVOGNL 60
|
QY 61 ELYLPTNASLSFLQDIQEVQGVVLIHNRQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
|
DB 61 ELYLPTNASLSFLQDIQEVQGVVLIHNRQVPLQRLIRVRGTOLFEDNYALAVLDNG 120
|
QY 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQORNPOLCYQDTILMKDIFHKNNOLA 180
|
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILMKDIFHKNNOLA 180
|
QY 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240
|
DB 181 LTLIDTNRGRACHPCSPCKGRCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240
|
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
|
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
|
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEDGTORCEKSKPCARVCYGLGMEHUREVRVTSAN 360
|
DB 301 YNYLSTDVGSCTLVCPHMQEVTAEDGTORCEKSKPCARVCYGLGMEHUREVRVTSAN 360
|
QY 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETIGLYLISAWPDSL 420
|
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETIGLYLISAWPDSL 420
|
QY 421 DLSVFONLQVIRGRILHNCAYSILTLQGLISWGLSLRELGLSLALIHHTHLCFVHTV 480
|
DB 421 DLSVFONLQVIRGRILHNCAYSILTLQGLISWGLSLRELGLSLALIHHTHLCFVHTV 480
|
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQCHQCHGWGPGTQCVNCSQFLRGQEC 540
|
DB 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQCHQCHGWGPGTQCVNCSQFLRGQEC 540
|
QY 541 VEECRVLOGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCVACAHYKDPFCVVARC 600
|
DB 541 VEECRVLOGLPREYVNAHCLPCHPECQFQNGSVTCFGEADQCVACAHYKDPFCVVARC 600
|
QY 601 PSGVRPDLSYMFIWKPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVWG 660
|
DB 601 PSGVRPDLSYMFIWKPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVWG 660
|
QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNOQMILKETEL 720
|
DB 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNOQMILKETEL 720
|
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKEILDYVMAGVGP 780
|
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIIPVAIKVLRNTSPKANKEILDYVMAGVGP 780
|
QY 781 YVSRLLGICLTSTVQLVLTQMPYGCGLLDHVRENRGRGLSGDQLLNWCQIAKGNMYSLEDVR 840
|
DB 781 YVSRLLGICLTSTVQLVLTQMPYGCGLLDHVRENRGRGLSGDQLLNWCQIAKGNMYSLEDVR 840
|
QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900
|

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Qy 961 IDSECRPRFRELSEFSRMDRDPORFVVIQNEIDLGASPLDSTFYRSILLEDGMDLVA 1020
Db 961 IDSECRPRFRELSEFSRMDRDPORFVVIQNEIDLGASPLDSTFYRSILLEDGMDLVA 1020
Qy 1021 EYLVPQGFPCPDPAACAGMVRHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAACAGMVRHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.
AC AAR39568;
XX
XX
DT 07-FEB-1994 (first entry)
DE Sequence of c-erbB-2 tumour antigen.
XX
XX Tumour antigen; c-erbB-2; glycoprotein.
KW Homo sapiens.
OS
XX
XX W09316185-A.
XX
XX 19-AUG-1993.
XX
XX 05-FEB-1993; 93WO-US01055.
XX
XX 06-FEB-1992; 92US-0831967.
XX
XX (CETU) CETUS ONCOLOGY CORP.
XX (CREA-) CREATIVE BIOMOLECULES INC.
XX
XX Houston LL, Huston JS, Oppermann H, Ring DB;
XX
XX WPI; 1993-272889/34.
XX N-PSDB; AAQ46083.
XX
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX
XX Disclosure; pages 48-54; 87pp; English.
XX
XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 98.2%; Score 6687; DB 14; Length 1433;
Best Local Similarity 98.3%; Pred. No. 0;

Matches 1234; Conservative 5; Mismatches 16; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASTQVCTGDMKURLPASPETHLDMLRHLVQGCVOVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASTQVCTGDMKURLPASPETHLDMLRHLVQGCVOVQGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVQGYVLAHNOVROVPLQRLIRVGTQFLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIQEVQGYVLAHNOVROVPLQRLIRVGTQFLFEDNYALAVLDNG 120
Qy 121 DPLNNQYIKANSKFIGITELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSESDCQSLTRTVGAGGCARCKGPLTDCCHQEC 240
Db 181 LTLIDTNRSRACHPCSPMKGSRGWSESDCQSLTRTVGAGGCARCKGPLTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVSGCTLVCPLHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEETCYLYLISAWPDSL 420
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEHLQVFTLEETCYLYLISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRI LHNGAYSILTLOGLGISWLGSLRLSELGSLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRI LHNGAYSILTLOGLGISWLGSLRLSELGSLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCGWGPPTOCVNCQSLRGOEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCGWGPPTOCVNCQSLRGOEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVAVRC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVAVRC 600
Qy 601 PSGVKPDLSPYMPKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSPYMPKWPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660
Qy 661 ILLVWLVGVVFGILIKRQOKIRKYMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVWLVGVVFGILIKRQOKIRKYMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWI PDGENVKI PVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWI PDGENVKI PVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRGSLDNLNWCQIAKGSYLEVDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRGSLDNLNWCQIAKGSYLEVDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMIWVKWM 960
Qy 961 IDSECRPRFRELSEFSRMDRDPORFVVIQNEIDLGASPLDSTFYRSILLEDGMDLVA 1020
Db 961 IDSECRPRFRELSEFSRMDRDPORFVVIQNEIDLGASPLDSTFYRSILLEDGMDLVA 1020
Qy 1021 EYLVPQGFPCPDPAACAGMVRHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
Db 1021 EYLVPQGFPCPDPAACAGMVRHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAKGLOSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255

Search completed: July 22, 2003, 08:40:28
Job time : 43.9774 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-149-163-12
Perfect score: 6810
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTFPTAENPEYLGLDVPV 1255
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 9613422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6734	98.9	1255	1 A24571	protein-tyrosine k
2	5935	87.2	1260	1 TVRTNU	protein-tyrosine k
3	5931.5	87.1	1254	2 I48161	p-185 precursor -
4	3160	46.4	1210	1 GQHUE	epidermal growth f
5	3135	46.0	1210	2 AS3183	epidermal growth f
6	3109.5	45.7	1223	1 TVCHLV	epidermal growth f
7	2975.5	43.7	1308	2 A47253	epidermal growth f
8	2684	39.4	1166	1 S06142	protein-tyrosine k
9	2423.5	35.6	1342	2 A36223	kinase-related tra
10	2339.5	34.4	1339	2 JCA387	epidermal growth f
11	1766.5	25.9	698	1 TVPVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUOH	protein-tyrosine k
13	1647	24.2	544	2 S35745	protein-tyrosine k
14	1645.5	24.2	1330	1 GQFFE	protein-tyrosine k
15	1640	24.1	545	2 S00727	epidermal growth f
16	1623	23.8	540	2 B44776	kinase-related tra
17	1621	23.8	540	1 TVPVEB	protein-tyrosine k
18	1530	22.5	644	2 A36325	protein-tyrosine k
19	1297	19.0	1323	2 E88257	epidermal growth f
20	1297	19.0	1374	2 S70712	protein let-23 (im
21	1208	17.7	1369	2 S70713	protein-tyrosine k
22	1162	17.1	1717	1 A45558	epidermal growth f
23	1141	16.8	527	2 A42032	epidermal growth f
24	990.5	14.5	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	725	10.6	1363	2 T43220	protein-tyrosine k
28	713	10.5	1382	1 INHUR	insulin-like growt
29	705.5	10.4	1372	2 A34157	insulin receptor p

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erbB-2 protein precursor; kinase-related transforming protein e
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; B44188; 159509; 157622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198
R:Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282
R:Cousens, L.; Yang-Peng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181; PMID:2999374
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'PALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:G183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: 159509; MUID:85272597; PMID:2992089
A:Accession: 159509
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: 157622; MUID:87286898; PMID:3039351
A:Accession: 157622
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

insulin receptor p
insulin-like growt
insulin receptor-r
insulin receptor-r
protein-tyrosine k
insulin-like growt
insulin-like growt
insulin-like growt
insulin receptor -
insulin receptor -
insulin receptor (
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k

30 705 10.4 1383 2 A36080
31 702 10.3 1607 2 T43212
32 698.5 10.3 1300 2 A36502
33 684 10.0 1268 2 B36502
34 669 9.8 1477 2 T18534
35 650 9.5 1367 1 IGHUR1
36 639 9.4 1371 2 A33837
37 634 9.3 1390 2 T30346
38 618.5 9.1 2148 1 A56081
39 615 9.0 2101 2 S57245
40 591.5 8.7 952 2 I50612
41 591 8.7 987 2 A54092
42 589.5 8.7 977 2 S49004
43 588 8.6 1114 1 S05582
44 586 8.6 1091 2 S33596
45 584.5 8.6 976 2 A36355

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAAS8637.1; PID:g553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase

F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <BE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.98; Score 6734; DB 1; Length 1255;
Best Local Similarity 98.88; Pred. No. 7.4e-268;
Matches 1240; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASQTGCTGDKMLRASPETHLDMRLRHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQTGCTGDKMLRASPETHLDMRLRHLYQGCVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGVYLIAHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELYLPTNASLSFLQDIQEVQGVYLIAHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILKWDQYIKANSKF 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILKWDQYIKANSKF 180
QY 181 IGTENLRSRACHPCSPMKGSRGWSESSDCQSLTRTVCAAGCARCKGLPTDCHEQC 240
DB 181 LTLLDTRSRACHPCSPMKGSRGWSESSDCQSLTRTVCAAGCARCKGLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
DB 301 YNYLSTDVGSCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
QY 361 IQEFAGCKKIFGSLAFPLPESFDGPASNTAPLOEQLVQFETLEEITGYLIYSAMPDLSL 420
DB 361 IQEFAGCKKIFGSLAFPLPESFDGPASNTAPLOEQLVQFETLEEITGYLIYSAMPDLSL 420
QY 421 DLSVFQNLQVIRGRIHLHNGAYSITLQGLISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRIHLHNGAYSITLQGLISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
QY 481 PWDLFRNPHOALLHTANRPEDSCVGEGLACHOLCARGHCWGPQTQVCNCSOFLRGQEC 540
DB 481 PWDLFRNPHOALLHTANRPEDSCVGEGLACHOLCARGHCWGPQTQVCNCSOFLRGQEC 540
QY 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600
QY 601 PSQVKPDLVSMPTIWKFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTSISAVVG 660
|||||

DB 601 PSQVKPDLVSMPTIWKFPDEEGACQPCPINCTHSCVDLDDKGCQPAEQRASPLTSISAVVG 660
QY 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLQSTELVEPLTPSGAMPNQAMRLKETEL 720
DB 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLQSTELVEPLTPSGAMPNQAMRLKETEL 720
QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
DB 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVLTQMLPYGCLLDHVRENRGLGSDQLLNCWCQIAKGSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVLTQMLPYGCLLDHVRENRGLGSDQLLNCWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRRFT 900
DB 841 LVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILRRRFT 900
QY 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
DB 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDSECRPRRELVSFESRWARDPQRFVITQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
DB 961 IDSECRPRRELVSFESRWARDPQRFVITQNEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020
QY 1021 EYVLVPOQGFCCPDPAAGAGMVHRRSSSTSGGDLTLGLEPSEEEAPRSLAPSEG 1080
DB 1021 EYVLVPOQGFCCPDPAAGAGMVHRRSSSTSGGDLTLGLEPSEEEAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140
QY 1141 NQPDVVRPQSPREGPLPAARPAAGATLERAKTLSPGKNGVYKQVFAFGGAVENPEYLTPO 1200
DB 1141 NQPDVVRPQSPREGPLPAARPAAGATLERAKTLSPGKNGVYKQVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYGLDVPV 1255
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohe

Carcinogenesis 12, 1973-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals r

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663; 'V', 665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

```
Query Match      87.1%; Score 5935; DB 1; Length 1260;
Best Local Similarity 87.1%; Pred. No. 3e-235;
Matches 1095; Conservative 51; Mismatches 109; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPGCAASTQVCTGTDMLRLPASPEHDLMLRHLVYGCQVVOGNL 60
Db 4 MELAAWCWGLLLALLPGIAGTQVCTGTDMLRLPASPEHDLMLRHLVYGCQVVOGNL 63
Qy 61 ELTYLPTNASISFLQDIOEVQGYVLIHNOVQVPLQRLRIVRGTLFEDNVALAVLNG 120
Db 64 ELTVVPANASISFLQDIOEVQGYMLIANQVQVPLQRLRIVRGTLFEDKVALAVLNR 123
Qy 121 DPLNNTTPTV-GASPGGLRELQLRLSLTEILKGGVLIQVLPQCYQDTILWKDQYIKANSK 179
Db 124 DPQDNVAASTPGRTPGRLQLRLSLTEILKGGVLIQVLPQCYQDMVLWKDQVPRKNQL 183
Qy 180 FIGITELNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVACGCARCKGPLPTDCCHQ 239
Db 184 APVDIDTNRSRACPPCAPCKDNHCWGESPEDCILTCTICTSGCARCKGRPLPTDCCHQ 243
Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
Db 244 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTC 303
Qy 300 PNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARCYGLGMEHLREVRATSA 359
Db 304 PNYLSTEVGSCTLVCPNNQNEVTAEADGTQRCCKSPCARCYGLGMEHLRGARITSD 363
Qy 360 NIQFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLEEITGYLIYISAWPDL 419
Db 364 NVQFQDCKKIFGSLAFIPESFDGDPSSGIAPLQEQLOVFETLEEITGYLIYISAWPDL 423
Qy 420 PDLVSFQNLQVIRGLHNGAYSLTLOGISLWGLRLSRLRELGLSLIHHNTHLCFVHT 479
Db 424 RDLVSFQNLRIIRGLHDGAYSLTLOGISLWGLRLSRLRELGLSLIHRNAHLCFVHT 483
Qy 480 VPWDLFRNPHOALLHTANPEDE-CVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQ 538
Db 484 VPWDLFRNPHOALLHSGNRPEEDLCVSSGLVCNSLCAHGCWGPPTQCVNCSHFLRGQ 543
Qy 539 ECVEECRVLOGLPREYVNRHCLPCHPECQPNQSVTCFGEADQCVAHYKDPFQVA 598
Db 544 ECVEECRVWKLPREYVSDKRLPCHPECQPNQSVTCFGEADQCAAHYKQSSSCVA 603
Qy 599 RCPGVKPDLSYMPITWKPPEEGACQPCPINCTHSCVDLDKGCPEAORASPLTSIVSAV 658
Db 604 RCPGVKPDLSYMPITWKPPEEGACQPCPINCTHSCVDLDKGCPEAORASPVTFIATV 663
Qy 659 VGILLVVLVGVFGILIKRQKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKET 718
Db 664 EGVLFLILVVGILIKRQKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKET 723
Qy 719 ELRKVKVLGSAFGTVYKGIWIPQENVKIIPVAIKVLENTSPKANKEILDEAYVMAGV 778
Db 724 ELRKVKVLGSAFGTVYKGIWIPQENVKIIPVAIKVLENTSPKANKEILDEAYVMAGV 783
Qy 779 SPYVSRLLIGICTSTVQLVTQMPYGCLLDVRNENRGLSGQDLLNWCQIAKMSYLED 838
Db 784 SPYVSRLLIGICTSTVQLVTQMPYGCLLDVRNENRGLSGQDLLNWCQIAKMSYLED 843
Qy 839 VRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 898
Db 844 VRLVHRLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 903
Qy 899 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDTVNIMVKC 958
```

RESULT 3

I48161

P-185 precursor - golden hamster

C:Species: Mesocricetus auratus (golden hamster)

C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999

C:Accession: I48161

R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika

Gene 140, 251-255, 1994

A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.

A:Reference number: I48161; MUID:94193007; PMID:7908275

A:Accession: I48161

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1254 <RES>

A:Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595

C:Genetics:

A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP

F:718-983/Domain: protein kinase homology <KIN>

F:726-734/Region: protein kinase ATP-binding motif

```
Query Match      87.1%; Score 5931.5; DB 2; Length 1254;
Best Local Similarity 86.9%; Pred. No. 4.2e-235;
Matches 1091; Conservative 59; Mismatches 104; Indels 1; Gaps 1;
```

```
Qy 1 MELAALCRWGLLLALLPGCAASTQVCTGTDMLRLPASPEHDLMLRHLVYGCQVVOGNL 60
Db 1 MELAAWCWGLLLALLPGASGTQVCTGTDMLRLPASPEHDLIRVHLYGCQVVOGNL 60
Qy 61 ELTYLPTNASISFLQDIOEVQGYVLIHNOVQVPLQRLRIVRGTLFEDNVALAVLNG 120
Db 61 ELTYLPANATLSFLQDIOEVQGYMLIAHQVHRVPLQRLRIVRGTLFEDKVALAVLNR 120
Qy 121 DPLNNTTPTVTCASPGGLRELQLRLSLTEILKGGVLIQVLPQCYQDTILWKDQYIKANSKF 180
Db 121 DPLDNVTATGRTPEGLRELQLRLSLTEILKGGVLIQVLPQCYQDTILWKDQVFRKNOLA 180
Qy 181 IGITELNRSRACHPCSPMKGSCRCWGESSEDCQSLTRTVACGCARCKGPLPTDCCHQ 240
Db 181 PVDIDTNRSRACPPCAPCKDNHCWGESPEDCILTCTIAPRAVPAARALPTDCCHQ 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEADGTQRCCKSPCARCYGLGMEHLREVRATSA 360
Db 301 YNYLSTEVGSCTLVCPPLNNOEVTAEADGTQRCCKSPCARCYGLGMEHLRGARITSA 360
```


A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A:Title: Functional independence of the epidermal growth factor receptor from a domain
A:Reference number: A3331; MUID:90003233; PMID:2790960
A:Contents: annotation; internalization signal
C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
C:Genetics:
A:Gene: GDB:EGFR
A:Cross-references: GDB:120610; OMIM:131550
A:Map position: 7p12.3-7p12.1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1210/Product: EGF receptor #status predicted <MAT>
F:25-645/Domain: extracellular #status predicted <EXT>
F:390-600/Domain: EGF receptor extracellular domain repeat <EE1>
F:669-1210/Domain: EGF receptor extracellular domain repeat <EE2>
F:646-668/Domain: transmembrane #status predicted <TM>
F:710-975/Domain: intracellular #status predicted <INT>
F:718-726/Region: protein kinase homology <KIN>
F:999-1046/Region: coated-pit mediated internalization signal
F:1047-1210/Region: inhibitory
F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F:745/Active site: Lys #status experimental

Query Match 46.4%; Score 3160; DB 1; Length 1210;
Best Local Similarity 49.8%; Pred. No. 5,4e-122;
Matches 631; Conservative 177; Mismatches 348; Indels 112; Gaps 23;

QY	11	LLALLPGAA--STQVCTGTDMKRLPASPETHDMLRHLVYOGCOVGNLELTPLTN	68
DB	14	LLAALCPASRALEKKVCQGTNKLTLQGTGFHFLSLQRMFNCEVILGNLEITYVQRN	73
QY	69	ASLSFLDDIQEVOGYVLIHNVQVPLQRLRVRGTQTFEDNYALAVLNDGDLNNTTP	128
DB	74	YDLSEFKTIQEVAGYVLIHNVQVPLQRLRVRGTQTFEDNYALAVLNDGDLNNTTP	126
QY	129	VTGASPGGLRELQLRLSLTEILKGGVLIQRNPOLCYOTILKQVTKANSKFIGITEL--	186
DB	127	---ANKTGLKELPMRLQLBILHGAVRFSNNPALCNVESIQWRD---IVSSDFLSNMSDF	180
QY	187	-NRSRACHPCSPMKGRSCWGSSEDCQSLRTVTCAGCA-RCKGLPLTDCHEQCAACG	244
DB	181	QNHLSGCRQCDPSCPNCSGWCAGEENCQKLTKIIAQCQSGRCGRKSPDCCHNQCAACG	240
QY	245	TGPKHSDCLACLHNHSGICELHCPALVTYNTDTESMPNPEGRYTFEGASCVTACPNYL	304
DB	241	TGPRSDCLVCRKFRDEATCKDTCPPLMLYNTTYQMDVNPPEGKVSFGATCVKRCPRNVY	300
QY	305	STDVGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSANIQEF	364
DB	301	VTDHGSCVRACADSYEM-EDGVKCKACEGCRKVCNGIGIGEFKDSLSINATNIKH	359
QY	365	ACKKIFGSLAPSPFDGDPASNTAPLOEQVFTLEETIGVLYISAWPDSLPDLSV	424
DB	360	KNCTISGDLHLPLVAFRGDSFTHTPPLDPOELDKVTEITGFLIQAWPENRTDLHA	419
QY	425	FONLQVIRGRIHLNGAYSITLQGLGSLGLSLRELGLALIHNNHLCFVHTVPWDO	484
DB	420	FENLEIIRGTQHQGFSLAVVSLNLSLGLSLKEISDGVIIISGNKRLCVANTINMKK	479
QY	485	LFRNPQALLHTANRDECEVCEGLACHOLCARGHCWGPQTQVCNCSOFLRGQCEVCB	544
DB	480	LFGTSGQTKIISNRGENSKATQVCHALCSPEGCGPEPRDCVSCRNVSRGRCVDKC	539
QY	545	RVLOGLPREYNARHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFFCVARCPGV	604
DB	540	KLLEGEPEFVENSECIOCHPECLPQAMNITCTGRGPONCIQCAHYIDGPHCVKTC	599
QY	605	KPDLSPMPKPFDEEGACQPCPINCSTHSCVDLDKGCAPAEQASPLTSIVSAVVG---	661

Db	600	MGENNTL-VNKYADAGHVCHLCHPNCTYCTGCTGEGCTNGPKIP--SIATGMVGALLL	656
QY	662	LLVWLVGVVFGILIKRQOKIRKYTMRRLLOETELVEPLTPSGAMPNQAMRILKETELR	721
Db	657	LLVVALGIG--LFMRRHIVKRTLRLLORELVEPLTPSGEAPNQALLRLKETEFK	713
QY	722	KVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSPY	781
Db	714	KIKVLGSAFGTVYKGLWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMASVDNPH	773
QY	782	VSRLLIGICLTSTVOLVTLMPYCLLDHVRNRLGSGDQLLNWCHQIAKMSYLEVRL	841
Db	774	VCRLLIGICLTSTVOLVTLMPYCLLDHVRNRLGSGDQLLNWCHQIAKMSYLEVRL	833
QY	842	VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMMALESILRRFFTH	901
Db	834	VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMMALESILRRFFTH	893
QY	902	QSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTIDVYIMVAKCMI	961
Db	894	QSDVMSYGVTVWELMTFGSKPYDGIPAREIPDLLEKGERLPQPPICTIDVYIMVAKCMI	953
QY	962	DSECPREPRELYSESRMARDPQRFVIO-NEDLGPASPLDSTFFYRSLLEDGMDLVDA	1020
Db	954	DADSRPKFRELIEFSKWARDPQRYLVICQDERMHLPSPTDSNFYRALMDEEDMDVDVA	1013
QY	1021	EBYLVPOQGFCCPDAPGAGGVHHRHSSTRSGGDLTLGLERSEEAAPRSLAPSEG	1080
Db	1014	DYLIPLQOQGF-----SSPSTRPLLSLS	1039
QY	1081	AGSDVFDGLGMAAKGLQLSPHPSQRYSEDPTVPLPSET--DGVVAPLTCSPQE	1138
Db	1040	ATSN--NSTVACIDRNLQSCPIKEDSFLQRYSSDPTGALTEDSIDTFL-----PVPE	1091
QY	1139	YVQPDVPRQPPSPREGPLPAARPAGATLAKTLSPGNGVGVKQVAFGGAVENPEYL-	1197
Db	1092	YINQ--SVPKRPAQSVQNPVYHQNPLN-----APSRDPYQD--PHSTAVGNPEYL	1140
QY	1198	TPQGGAAQPHPPAFSPAFDNLYWDQ-----DP-----PERGAPSPFKGTP	1241
Db	1141	TVQ-----PTCVNSTFDSPAHWAQSGHQISLDNPDYQODFFPKAKPNIGI	1190
QY	1242	TAENPEYL 1249	
Db	1191	TAENAEYL 1198	

RESULT 5
A53183
epidermal growth factor receptor precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C:Accession: A5318; A43818; S24942; A28941; S45325; I49643
R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Jenkins, N.A.
Genes Dev. 8, 399-413, 1994
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A:Reference number: A53183; MUID:94170986; PMID:8125255
A:Accession: A53183
A:Molecule type: mRNA
A:Residues: 1-1210 <LUE>
A:Cross-references: GB:U03425
R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
A:Reference number: A43818; MUID:91232866; PMID:2030916
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eisinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R.Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A06643; MUID:85228222; PMID:2988784
A:Accession: A06643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <EXT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.7%; Score 3109.5; DB 1; Length 1223;
Best Local Similarity 48.6%; Pred. No. 6.3e-120;
Matches 630; Conservative 175; Mismatches 347; Indels 145; Gaps 25;

QY 8 RMGLLALLPPGAA-----STQVCTGTDMKRLPASPETHLDMLRHLHYOGQVQGNLE 61
DB 13 RGAALVLLVLLGVALCSAVEEKVCGQTNNKLTQLGHVEDHTSLQRMNNECVLSNLE 72
QY 62 LTYLPNALSFLQDIQEOGVYLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNGD 121
DB 73 ITYVEHNRDLTKTIQEVAGYVLIANMVDVPLENLQIRGNVLYDNSFALVLSNVH 132
QY 122 PLNNITPVGASPGRLRELQSLRSLTEILKGGVLIQRNPOLCYQDTILWKDQYIKANSKFI 181
DB 133 -MNKTQ-----GLRELPLKRLSELNGGVKISNNPKLCNMMDTVLMND-IIDTSRKPL 182
QY 182 GITEL-NRSRACHPCSPMCKSGRSGESSEDCQSLTRTYCAGCA-RCKGPLPTDCCHQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCOTLTKVICAQCGSRCKRVPSCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRESDCLACRKRDDATCKDTCPLVLYNPTTYQMDVNPPEGKYSFGATCVR 302
QY 300 PYNLYLTDVGSCTLVCPHLNQBVTAEQTCREKSKPCARVCYGLMGHEHLREVAVTSA 359
DB 303 PHNYVYTDHGSCVRSNTDTYEV-BENGVRKCKCDGLCSKVCNGIGIGELGKILSINAT 361
QY 360 NIOEFAGCKKIFGSLAFPESEFDGPASNTAPLOEQLOVFTLEETITGLYLIASWPDSL 419
DB 362 NIDSFKNCTKINGDVSILPVALFGDAFTKTLPLDKLDFVTVKEISGFLLIQAWPDNA 421
QY 420 PDLVSFQNLQVIRGRILHNGAYSLTQGLIGISWGLRLSRLRELSGLALIHNNHLCFVHT 479
DB 422 TDLYAFENLEIIRGTRTKHQGYSLAVNLKIQSLGLRSLKEISDGDIAIMKNKILCYADT 481
QY 480 VPMDOLERNPHQALLHTANPEDECVGEGELACHQICARGHCWGPQPTQVNCQSLRQGE 539
DB 482 MNWRSLFATQSQTKI1QNRNKNCDTADRHVCDPLCSVDGCGWGPFGFCFCRFFSRQKE 541
QY 540 CVEECRVLQGLPREYVYNARHCLPCHPECPQNG---SVTCFGPEADQCVACAHYKDPFC 596

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C:Accession: A47253

R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.;

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A:Title: Ligand-specific activation of HER4/pl80erbB4, a fourth member of the epiderm

A:Reference number: A47253; MUID:93189574; PMID:8383326

A:Accession: A47253

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1308 <PLO>

A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360

A>Note: sequence extracted from NCBI backbone (NCBIP:126842)

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor

F:716-981/Domain: protein kinase homology <KIN>

F:724-732/Region: protein kinase ATP-binding motif

Query Match		43.7%	Score 2975.5;	DB 2;	Length 1308;
Best Local Similarity		45.2%	Pred. No. 2e-114;		
Matches		609;	Conservative 185;	Mismatches 379;	Indels 173; Gaps 28;
QY	9	WGLLLALLPGAA-----	STOVCTGTDKMLRPLASPETHLDMLRHLRYOGCOVVOGNEULTY	64	
Db	8	WVWSLLVAAGTVQPSDSQVCAGTENKLSLSLEQOYRALRYKYNCEVVMGNLEITS	67		
QY	65	LPTNASLFLQDIQEVGYVLI	IAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNGDPLN	124	
Db	68	IEHRDLSFLRSVREVTGYVLVALNQFYLPLENLRIIRGTLYKLYEDRYALAI	FLNRYKDG	127	
QY	125	NTTPVTGASPGGRLQLRSLTEILKGGVLIQRNPOLCYQDTILWKDQYIKANSKFIGIT	184		
Db	128	NF-----	GLQELGLKNLTEILNGVGVVDQNKFLCYADTIHQWDIVRNWPSNLTIV	178	
QY	185	ELNRSRACHPCSPMKGSRGWSSESDECSLRTVTCAGGC-ARCKGRLPTDCHEQCAAG	243		
Db	179	STNGSGCGRCHKSCTG-RCMGPTENHCOTLTRTVCAEQCDGRCYGPVYSDCHRECAAG	237		
QY	244	CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFEGASCVTACPNY	303		
Db	238	CSGPKOTDCFACHNFNDSGACVTCQPOTFYNPTTFOLEHNFNAKYTYGAFCKKCPHNF	297		
QY	304	LSTDVGSCTLVCLPHNOEVTAEBDQRCCEKSRPCARVCYGLGMEHLREVRVTSANIOE	363		
Db	298	V-VDSSSCVACRCPSSKMEV-EENGIKWCKPCTDICPKACDGI	GTGSLMSAQTVDSNIDK	355	
QY	364	FACKKIFGSLAFPEFSDGDPASNTAPLOEQLOVFETLEBETGYIYISAWPDSL	PDLS	423	
Db	356	FINCTKINGNLIFLVTGIHGDYNAIBADPEKLNVRTVREITFGLNIQSWPPNMTDFS	415		
QY	424	VFNQLQVIRGRILHNGAYSITLQGLGISWLGSLRSLGSLALIHNTHL	CFVHTVPWD	483	
Db	416	VFNLTVIGRVLVYSGLSLILKQOGITSLOFQSLKEISAGNIYITD	NSLCLYHINWT	475	
QY	484	QLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRQECVEE	543		
Db	476	TLFSTINQIRVIRDNRAENCAETAGMVCNHLCSDDGCGWGPDPQCLSCRRFSRGRICIES	535		
QY	544	CRVLOGLPREYVNAHCLPCHPECOP-QNGSVTCFGEADQCACAHYKDPFPCVACRPS	602		
Db	536	CNLYDGEFRFENGISICEVCDQCEKMEDEGLLTCHGPGDPNCTKCSHFKDGPNCVEKCPD	595		
QY	603	GVKPDLSYMPIWPFPPDEGACQPCPINCTHSCSDVLDKGC-----	PAEQRASPL	651	
Db	596	GLOGANSF--IFKVADPDRECHCPNCTQCGNPTSHDCIYVPWCHSLTPQHAR-TPL	652		
QY	652	TSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOA	710		
Db	653	--IAAGVIGGLFVLVGLTFVAVVRRKSIK-KKRALRRFL-ETELVEPLTPSGTAPNOA	708		
QY	711	QMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	770		
Db	709	QLRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE	768		
QY	771	AYVMAGVGSVPSRLIGLICLTSTVOLVTQLMPYGCILLDHVRENKRGSLGSDLLNWCQIA	830		
Db	769	ALIMASNDHPLVRLVGLVCLSPITQIVTQLMPHGCCLLEYVHEKDNIGSOLLNWCQIA	828		
QY	831	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPKIMMA	890		
Db	829	KGMYLEERLVRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGGKVPKIMMA	888		
QY	891	LESILRRFRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	950		
Db	889	LECILHRYKFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID	948		
QY	951	VYIMVVKCMIIDSECRPRFRELSEFSRMDARDPQRFVVIQNEO-LGPASPLDSTFVRSLL	1009		
Db	949	VYIMVVKCMIIDADSRPKFELAAEFSSRMDARDPQRYLVIOGDDRMKLPSPNDSKFFQNL	1008		
QY	1010	ZDDMGDLVDAEYLVLPQOQFFCPDPPAGAGGVHHRHRSSTRSGGGDLTLGLEPSEEE	1069		

Db	1009	DEEDLEMDMAEYLVLP-QAFNIPPP-----IYTSRARIDSNRS-----EIGHSPPPAY	1056
QY	1070	APRS-----PLAP-SEGAGSDVFDGDLGMAAKGLQS	1100
Db	1057	TPMSGNQFYVRDGGFAAEQGSVVPYRAPTSTIPEAPVAQGATAEII	1116
QY	1101	LPHDPSPLQRYSESDTVPPLPS-----ETGCVAPLTCSPQPEYVNPQDVRPQPPSPR	1153
Db	1117	PHVQEDSSQRYVSADTVPFAPERSPRGELDEEGYMTFMRDKPKQEYLN	1167
QY	1154	EGPLPAARPAAGATLERAKTLPCKGVKVDVFAFGGAVENPEYLT	1212
Db	1168	ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAE	1199
QY	1213	-----FSPAFONLYYWDQDPPERGA--PPSTF	1237
Db	1200	DEYVNEPLYLNTFANTLGKAEYKKNILSMPEKAKAFDNP	1259
QY	1238	KGTP-----AENPEYL	1249
Db	1260	QEYSTKYFYKQNGRIRPIVAENPEYL	1285
RESULT 8			
S06142			
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish			
N;Alternate names: epidermal growth factor receptor homolog; kinase-related trans			
C;Species: Xiphophorus maculatus (southern platyfish)			
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000			
C;Accession: S06142; S13809			
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robe			
Nature 341, 415-421, 1989			
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-induc			
A;Reference number: S06142; MUID:90015140; PMID:2797166			
A;Accession: S06142			
A;Molecule type: DNA			
A;Residues: 1-1166 <WIT>			
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291			
R;Adam, D.; Maeueller, W.; Schartl, M.			
Oncogene 6, 73-80, 1991			
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in X			
A;Reference number: S13807; MUID:91125882; PMID:1846957			
A;Accession: S13809			
A;Status: preliminary; translation not shown			
A;Molecule type: DNA			
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>			
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285			
C;Genetics:			
A;Gene: mrk			
A;Map position: Y			
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1			
C;Superfamily: epidermal growth factor receptor; protein kinase homology			
F;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane prote			
F;1-25/Domain: signal sequence #status predicted <SIG>			
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <M			
F;707-972/Domain: protein kinase homology <KIN>			
F;715-723/Region: protein kinase ATP-binding motif			
Query Match			
Best Local Similarity 45.2%; Pred. No. 1.4e-102;			
Matches 573; Conservative 165; Mismatches 391; Indels 138; Gaps 28;			
QY	4	AALCRGILLALLPPGAAT-----QVCTGTDKMLRPLASPETHLDMLRHLRYOGCQVOGN	59
Db	8	AALLQ--LLLVLSISRCCSTDPDRKVCQGTSNQMTM---LDNHYLKMKNKMYSGCNVYLEN	62
QY	60	LELTPLTNASLSFLQDIOEVQGYVLI	119
Db	63	LEITYTQENQDLSFLQDIOEVQGYVLI	122
QY	120	GDPLNNTTPTVTGASPGGLRELQLRSITEILKGGVLIQRNPOLCYQDTILWKDQYIKANSK	179

Db 123 YQK-NPSSP--DVYQVGLKQLQLSNLTETLSGGVKVSHNPLLCNVETINWMDIVDKTSNP 179
QY 180 FIGITELNRSRACHPCSPMCKGSRGWESSEDCQSLTRTVTCAGGC-ARCKGPLPTDCCH 238
Db 180 TWMILPHAFEROCOKCDHGCWNGSCWAPGPHGCHQKFTLLCAEQCNRRCRGPKPIDCNE 239
QY 239 QCAAGCTGPKHSDCLAHFNHSGICELHCPALVTVNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDNDDGTCKTCTPPKIYDIVSHQVVDVNPNIKYTFGAACVKE 299
QY 299 CPVNYLSTDVGSCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVTS 358
Db 300 CRSNVVUTE-GACVRSASAGMLEVD-ENGKRSCKPDGVCVKCDIGIGISLNTIAVNS 357
QY 359 ANIQEFAGCKIFGSLAFIPESPDGDPASNTAPLQPEQLQVFTLEETIGLYLISAWPDS 418
Db 358 TNIRSFNSCTKINGDIILNRNSPEGDPHYKIGTMDPEHLWNLTVTKEITGYLVIMWPN 417
QY 419 LPDLSVFONLVRIGRILHNGAYS-LTLOGLGISWGLSLRSLRGSLALIHNTLFCV 477
Db 418 MTSLSVFQNLRIIRGTTTSRSGFSFVVQVVRHLQWLGRLSLKSEVSNVILKNTLQRYA 477
QY 478 HTVPMDQLFRNPHQALLHTANRPEDECVEGLACHOLCARGHCWGPQTQVCNCSQFLRG 537
Db 478 NTINRRLFRSEDQSIYDART-----ENQTCNNECEDGCM-PGPTWCSCLVHVRG 529
QY 538 QBCVEECRVLQGLPRYVNAHCLPCHPECQPNGSVTCFPGPADQCVACAHYKDPFCV 597
Db 530 GRCVASCNLLQGEPREAQVDRGVCQCHQCLVQDLSLTCYGPANCKSAHFQDPQCI 589
QY 598 ARCPSCVKPDLWSMPTWKPEDEGACQPCINCTHSCVDLDDKGCFAEQBASPLTSIVA 657
Db 590 PRCPHGIDGDTL-LWKYADRMGQCQPCNQNCQTCGSGPGLSGRCD-IVSHSLAVGL 647
QY 658 VVGILLVVVLGVVFGVILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNOAQMILKE 717
Db 648 VSGLLITVALLIVLLRRRIK-RKTRICLLQEKELVEPLTPSGQAPNQAFRLKE 706
QY 718 TELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYWAGV 777
Db 707 TEFKKORVLGSGAFGTYYKGLWNPDCENIRIPVAIKVLRNTSPKYNQEVLEAYWASV 766
QY 778 GSPYVSRLIGLICLTSTVOLVTQMLPYGCLLDVHRENRGLSGODLLNWCQIAKMSYLE 837
Db 767 DHPHVCRLGICLTSAVOLVTQMLPYGCLLDVYRQHERICGQWLNNWCQIAKGNYLE 826
QY 838 DVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETVHADGGKVPKMWALESLRR 897
Db 827 ERHLVHRDLAARNVLKSNHVKITDFGLSKLLTADKEYQADGGKVPKMWALESLQW 886
QY 898 RTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVK 957
Db 887 TYTHQSDWSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPPOPICTIEVYMLK 946
QY 958 CMWIDSECPRELVESESRWARDQRFVITQNEDLGPASPLDSTFYRSLLDDMDGL 1017
Db 947 CMWIDSSPRFRELVEGFSQMDPSRYLVLTQ-----NLPSLSDRFLSRLLSSDD--DV 1001
QY 1018 VDAEYLVYPOQGFCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEEARPSPLAP 1077
Db 1002 VDAEYLLPYKXI-----NRQGS-----EPCIP 1024
QY 1078 SEGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDPV-PLPSETDGYVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSITLRNISDPTQNALEKLDGH-----1055
QY 1137 PEYVNPQVRPOP-----PSPRE-----GLP-AARPAGATLERAKTLSPGKNGVKD 1183
Db 1056 -EYVNPQGETSSRLSDIYNPYEDLTDCGWGPVSLSSQEAETNFRPEYVLTNTQNSL----1111
QY 1184 VFAFGAVENPEYLTPOGGAAPQPHPPPAFSAFNLYYWDQDPPERGAPPSTFKGTPTA 1243
Db 1112 PLVSSGSDDDPY-----QAG-----YQAAF-----LPQTGALTONGMFLPAA 1149

QY 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223

kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human

C;Species: Homo sapiens (man)

C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C;Accession: A36223; 159164

R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epiderma

A;Reference number: A36223; MUID:90083234; PMID:2687875

A;Accession: A36223

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1342 <KRA>

A;Cross-references: GB:M29366

R;Procman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro,

proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A;Title: Molecular cloning and expression of another epidermal growth factor receptor

A;Reference number: 159164; MUID:90311312; PMID:2164210

A;Accession: 159164

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A;Cross-references: GB:M34309; NID:g183990; PIDN:AAA35979.1; PID:g306841

C;Genetics:

A;Gene: GDB:ERBB3; HER3

A;Cross-references: GDB:119880; OMIM:190151

A;Map position: 12q13-12q13

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hon

C;Keywords: ATP; phosphotransferase

F;707-972/Domain: protein kinase homology <KIN>

F;715-723/Region: protein kinase ATP-binding motif

Query Match 35.6%; Score 2423.5; DB 2; Length 1342;

Best Local Similarity 40.7%; Pred. No. 6.9e-92;

Matches 533; Conservative 191; Mismatches 458; Indels 129; Gaps 32;

QY 10 GLLALLPPGAA--STQVCTGTDMLRLPASPEHLMLRLHLYQGQVVOGNLELTVLPT 67
Db 11 GLLFSLARGSEVNSQAVCPGLTGLSVTGDENQVOTLYKLYERCEVWGNLEIVLTGH 70
QY 68 NASLFLQDIQEVQGVYLIHQNVRQVPLQRIVRGTQLFEDNYALAVLNDGDLNMTT 127
Db 71 NADLSFLQIREVTGYVLVAMNEFTLPLNLRVVRGTQVYDCKFAIFVM-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSLTEILKGGVLIQIORNPOLCYQDITLWKDQYIKANSKFIGITELN 187
Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKDKLCHMDTIDWRDADAE---IVVKD 178
QY 188 RSRACHPCSPMCKGSRGWESSEDCQSLTRTVTCAGGC-ARCKGPLPTDCCHCCAGCTG 246
Db 179 NGRSCPPCHEVCKG-RWGPGEEDCQTLTKTICAPQCNHCHGCPNPQCHDCGACGCG 237
QY 247 PKHSDCLACLFHNSGICELHCPALVTVNTDTFESMPNPEGRYTFGASCVTAOPYNLST 306
Db 238 PQDTCFACRHNDSGACVPRCPQPLVYNKLTQLEPNPHTTKYQYGVGVASCPHNFV-V 296
QY 307 DVGSCCTLVCPHLNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVTSANIQEFAG 366
Db 297 DQTSVCVRACPPDKMEVD-KNGLKMCPECGGLCPKACEGTGSG--SRFQTVSSNIDGFVN 353
QY 367 CKKIFGSLAFIPESPDGDPASNTAPLQPEQLQVFTLEETIGLYLISAWPDSLPDLISVQ 426
Db 354 CTKILGNLDFLTUNGDPWHKIPALDPEKLVNFTVREITGYLNTQSPPHMHNFVSFVS 413
QY 427 NLQVIRGRILHNGAYS-LTLOGLGISWGLSLRSLRGSLALIHNTLHNTLFCVHTVPDOL 485

```
Db 414 NLTTIGRSLYNRGFSLLIMKNLNVTSLSGFRSLKEISAGRIYISANRQLCVHHSLNWTKV 473
Qy 486 FRNPQHALLHTA-NRPDECEVGEGLACHOLCARGHCWGCPPTQCVNCSCQFLRGQECVSEC 544
Db 474 LRGPTEERLDIKHNRPRDCVAEGKVCDPICSSGGCWGPGQCLSCRNYSRSGVCVTHC 533
Qy 545 RVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHAYKDPFPFCVACRPSGV 604
Db 534 NFLNGEPREFAHEAEFSCHEPCOMEGTATCNGSGSDTCAQCAHFRDGHPCVSSCPGV 593
Qy 605 KPDLSPYPIWKFPDEBAGAQCPINCTHSCVDLDDKGCAPAEORA-----SPLTSIVSAVG 660
Db 594 LG--AKGPIYKYPDVQNECRPCHENCCTQGCKGPELQCLGQTLVLIGKTHLTALTVIAG 651
Qy 661 ILLVVLGVVFGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNOAQWRILKETE 719
Db 652 --LVVIFMILGGTFLYWRGRRIQNKRAMRYLGERGESIEPLDPS-EKANKVLARIFKETE 708
Qy 720 LRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLRENTSPKANKELIDELVYVAGVGS 779
Db 709 LRKLVLGSGVGTGVHKGVWIPGESIKIPVICIKVIEDKSGRQSFOAVTDHMLAIGSLDH 768
Qy 780 PYVSRLLGICLTSTVQLVTLQMPYGCULLDVHRENRRGLSQDLLLLNWCMIKAGNSYLEDV 839
Db 769 AHIVRLGLCPGSSQLVTVQLPLGSLLDHVQRHGAIGPQLLLNMGVQIAKGMYYLEEH 828
Qy 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRRF 899
Db 829 GMVHRNLAARNVLKSPQOVADFGVADLLPPDDKQLLYSEAKTPIKMWALESIHFGKY 888
Qy 900 THOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCW 959
Db 889 THOSDVMSYGVTVWELMTFCAEPYAGRLAEVPLLEKGERLAQPOICTIDVYIMVKCW 948
Qy 960 MIDSECPRELVSERFMRARDQRFVITQNEDLGPA---SPLDSTFYRSLLEDDMDGD 1016
Db 949 MIDENIRPTEKELANEFTWARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEVEELEP 1007
Qy 1017 LVDAEYLVPOQGFCDPAPAGAGVHRRHSSTRSGGDLTLGLPEP-SEBEAPRSPL 1075
Db 1008 ELDDLDEAED-----NLATTLSALSPLVGTINRPGSOSLL 1048
Qy 1076 APSEGAGSDVFDGLGMAAGLQSLPHD-PSPLQRYSEDPTVPLP-----SETDGYV 1128
Db 1049 SPSSGY-MPMNQNLGSCQESAVSGSERCPRPVSLH-----PMRPGCLASSSSGHV 1101
Qy 1129 A-----PLTCSPOPE-----YVNPQDVRPQPPSPREP-----L 1157
Db 1102 TGSEAELOEKVSMCRSRSRSPRGRGDSAYHSQRHSLLTPTVPLSPGLEEEDVNGYVM 1161
Qy 1158 PAARPAGATLERAKTLP-GKNGV-----KDVFAFGGAVENPEYILTPQGAAPOPHP 1210
Db 1162 PDTHLKTPSRSREGLTSSVGLSSVLGTEEDED-----EYEYWNRRRRHSR-PHP 1212
Qy 1211 PAFSPAFDNLIYYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
Db 1213 RPSSEELGVEYMDVGDLSASLGSTQCPLHPVPIPTAGTTDDEDEYVM 1263
```

RESULT 10

```
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>
```

```
A;Cross-references: GB:U29339; NID:g915389; PID:g915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GTT for resic
C;Comment: This protein is a functional heregulin receptor that transduces signals to
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hon
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)
```

```
Query Match 34.4%; Score 2339.5; DB 2; Length 1339;
Best Local Similarity 40.8%; Pred. No. 1.9e-88;
Matches 523; Conservative 172; Mismatches 433; Indels 155; Gaps 35;
```

```
Qy 3 LAALCRWGLLLALLPGAA---STQVCTGTDMLRLPLASPETHLDMLRLHLYQGCVVQGN 59
Db 7 LQVLC-----FLLSLRGSEMGNSQAVCPGTUNGLSVTGDDADNQVQTLYKYEKCEVWNG 62
Qy 60 LEITYLPTNASLSFLQDIOEVQGVYLIAHNQVRQVLRIRVRGTQLFDNYALAVLDN 119
Db 63 LEIVLTCHNADLSFLQWIREVTAYLVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-- 120
Qy 120 GDLNNTPTVTGASPGGLRELRLSRLTEILKGGVLIOERNPOLCVQDITLWKDQVIKANSK 179
Db 121 ---LNYNT----NSSHALRQLKFTLQTEILSGGVYIEKDKLCHMDTIDWRD-IVRVRGA 172
Qy 180 FIGITELNRSRACHPCSPCKSGRSGESSEDQSLTRTVCAGC-ARCKGPLPTDCCHE 238
Db 173 EIVVK--NNGANCPCHCEVKCG-RCWGGPDDQILTKTICAPCNCRCGFGPNQCCHD 229
Qy 239 QCAAGCTGPKHSCLACLHFNHSGICELHCPALVTYNTDTFESNPNEGRYTFGASCVTA 298
Db 230 ECAGGCGPQDTCFACRRENDSGACVPRCPPELVYNKLTFOLEPNPHTKYQYGGVCVAS 289
Qy 299 CPNYLSTDVGSCTLVPLHNQVETADGTQRECKSKPCARVCYGLGMEHLREVRATVS 358
Db 290 CPNHFV-VDQTFVCRACPPDKMEVD-KHGLKMECPCGGLCPKACEGTSG--SRVQTVD 345
Qy 359 ANIQEPAGCKKIFGSLAFLPESFGDPSANTAPLOPELOVFFETLEEITGLYLTISAMPDS 418
Db 346 SNTDGFVNCTKIQLNGLDGLITGLNDPWHKIPALDPELVNVRTVREITGLNTLSQWPPH 405
Qy 419 LPDLSVFQNLQVIRGRILHNHAYS-LTLOGLGISWLSGLRSLRELGSGLALIHHTLFCV 477
Db 406 MHNFSVFNLTITIGRSLYNRGFSLLIMKNLNVTSLSGFRSLKEISAGRVYISANQQLCYH 465
Qy 478 HTVPWDLFRNPQHALLHTA-NRPDECEVGEGLACHOLCARGHCWGCPPTQCVNCSCQFLR 536
Db 466 HSLNWTLLRLLGPSEERLDIKYDRPLGELAEKGVKCDPLCSSGGCWGPGQCLSCRNYSR 525
Qy 537 GQSCVEECRVQLGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVCAHAYKDPFPFC 596
Db 526 EGVCTVHCNPLQGEPRFVHEAQCFSCHPECLPMEGTSTYNGSGSDACARCAHFRDGHPC 585
Qy 597 VARCPSGVKPDLSYMPITWKFPPDEEGACQPCPINCTHSC--VDLDDKGCAPAEORASPLTSI 654
Db 586 VNSCPHGILG--AKGPIYKYPDQAQNECRPCHENCCTQGCNGPELQDCLGQAEVLMSKPHLV 643
Qy 655 VSAVGIILLVVLGVVFGILIKRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNOAQWR 713
Db 644 IAVTVG--LAVILMLGGSFYWRGRRIQNKRAMRYLGERGESIEPLDPS-EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLRENTSPKANKELIDELVY 773
Db 701 IPKETELRKVLGSGVFGVTHVHGIWIPGESIKIPVICIKVIEDKSGRQSFOAVTDHMLA 760
Qy 774 MAGVSPYVSRLLIGICLTSTVQLVTLQMPYGCULLDVHRENRRGLSQDLLLLNWCMIKAGM 833
```

Db '761 VGSLDHAIHVRLGLCGSSQLQVTOYPLGSLDLDHVHQHRETLGQPLLNNWGVQIAKGM 820

QY 834 SYLEDVRLVHRDLAARNLVKSPNHWK1TDFGLARLLDDIDETEHADGKVPKWMMALES 893

Db 821 YYLESHSVHRDLALRNVLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWMMALES 880

QY 894 ILRRRRETHOSDVMSYGVTVWELMTGCAKYDGIPIAREIPDLLEKGERLPQPPICITIDVYM 953

Db 881 IHFGKYTHOSDVMSYGVTVWELMTGCAEYAGLRUAEIPDLLEKGERLAQPPICITIDVYM 940

QY 954 IMVKCWMIDSECRPRELVSFBSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDEDD 1013

Db 941 VMVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GEPTP--PAAPSVLTKE 997

QY 1014 MGLVDABEYLVPOQGFCCPDPAAGGVMVHRRHRSSTRSGGDLJTLGLEPSEE----- 1068

Db 998 L-----QEAELEPEL-----DLDLDEAEELGLATS 1023

QY 1069 -----EAPRSLAPSG-----AGSDVFDGLGMAKGLQSLPTHD 1105

Db 1024 LGSALSPTGTUTRFGSOSLLSPSGYMPNMQSSLGEACLSAVLGGREQFSRPISTLH- 1082

QY 1106 PSFLQRYSDPTVPLPSETDGYV---APL-----TC-----SQPE-----YVNPQDV 1145

Db 1083 PIPRCR-----PASESEGHVTGSEAELEQKVSVCRSRSRSPRPGRGDSAYHSQRHS 1135

QY 1146 RPQPSPREGP-----LPAARPAGATLERAKTLSP-GKNGV-----KDVFAF 1187

Db 1136 LLTPVTPLSPGLEBEDGNGYVMPDTHLRGASSREGTLSSVGLSSVLGTEDEDE---- 1191

QY 1188 GGAIVENPEYLTPOGGAAPOPHP 1210

Db 1192 ----BEYEYMNRRKRGSP-PRPP 1209

RESULT 11

TVFVLV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C:Species: avian leukosis virus, ALV

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999

C:Accession: B00643; A00643

R:Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1985

A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A:Reference number: A00643; MUID:85228222; PMID:2988784

A:Accession: B00643

A:Molecule type: mRNA

A:Residues: 1-698 <N1>

A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750

A:Note: In Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal

C:Comment: This protein is synthesized as a gag-env-erbB protein.

C:Genetics:

A:Gene: gag-env-erbB

A:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific

F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:17-59/Product: env protein (fragment) #status predicted <ENV>

F:160-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>

F:194-459/Domain: protein kinase homology <KIN>

F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;

Best Local Similarity 52.2%; Pred. No. 2.4e-65;

Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 578 GPEADQCVACAHYKDDPPFCVACRPSGVKPDLSYMPIWFPDDEGACQPCPINCTHSCVDL 637

Db 60 GP--DHCKMKCAHFIDGPHCVKACPAGVLGENDTL-VKMYADANAVCQLCHPNCTRGCKGP 116

QY 638 DDKGCFAFORASPLTSIVSAVV-GILLVVVLGWFGLIKRQKIRKYTWRRLLQETEL 696

Db 117 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKRTLRLQLQREL 172

QY 697 VEPLTPSGAMPNQAMRIILKETELRKVKVGLSGAFCTVYKGIWIPDGENVKIPIVAKVLR 756

Db 173 VEPLTPSGEAPNQAHILILKETEFKVKVGLSGAFCTVYKGLWIPGEKVKIPIVAKELR 232

QY 757 ENTPSKANKIILDEAYVMAGVGSPPVYSRLGLICLTSTVQLVTOLMYPYGCLLDHHVNRGR 816

Db 233 EATSPKANKIILDEAYVMASVDNPHVCRLLGLICLTSTVQLITOLMYPYGCLLDIYREHKON 292

QY 817 LGSODLLNCWMOIAKMSVLEDRVLVHRDLAARNVLKSPNHWK1TDFGLARLLDDIDETE 876

Db 293 IGSQVLLNCWQIAKGMNLEERLVRDLAARNVLVKTPOHVKIITDFGLAKLGLGADEKE 352

QY 877 YHADGKGVPIKWMMALESILRRRFTHQSDVMSYGVTVWELMTGCAKYDGIPIAREIPDLLE 936

Db 353 YHAEGKGVPIKWMMALESILHRIYTHQSDVMSYGVTVWELMTGSKPYDGIPIASEISSVLE 412

QY 937 KGERLPQPPICITIDVYMIMVKCWMIDSECRPRELVSFBSRMARDPQRFVVIQ-NEDLG 995

Db 413 KGERLPQPPICITIDVYMIMVKCWMIDADSRPKFRELIAEFSKWARDPPRYLVITQGDERMH 472

QY 996 PASPLDSTFYRSLLDEDDMDGLVDAEYLVPOQGFCCPDPAAGGVMVHRRHRSSTRSG 1055

Db 473 LPSPTDSKPYRTLMBEEDMEDIVDADEYLVPHQGF-----NSPST---- 513

QY 1056 CGDLTLGLEPSEEEAPRSP-----APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLQ 1110

Db 514 -----SRTPLLSSLSATSNNSATNCID-----RNGQCHPVRESFVQ 550

QY 1111 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLE 1168

Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNQ--LMPKPS----- 585

QY 1169 RAKTLSPGKNGVVKDVF-----AFGGAIVENPEYLTPOGGAAPOPHPPPAF 1213

Db 586 ----TAMVQNYNNISLTAISKLPMDSRKYQNSHSTAVDNPEYL-----NTNQSPLA 633

QY 1214 SPAFNLVYWDQ-----DPPE-----RGAPPSTFKGTPTAENPEYVGLDVP 1254

Db 634 KTVFESSPWYQSGNHQINLNDPVDYQDDFLPNETKPNGLLKVPAAENPEYLVRAAP 689

RESULT 12

TVYUHV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)

C:Species: avian erythroblastosis virus

C:Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C:Accession: A00644; A38022

R:Yanamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-78, 1983

A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fa

A:Reference number: A00644; MUID:84026539; PMID:6313229

A:Accession: A00644

A:Molecule type: DNA

A:Residues: 1-604 <YAM>

A:Cross-references: GB:K01216; NID:g209676; PIDN:AAA42400.1; PID:g209678

R:Debuire, B.; Henry, C.; Benaissea, M.; Bisette, G.; Claverie, J.M.; Saulle, S.; Marti

Science 224, 1456-1459, 1984

A:Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type

A:Reference number: A38022; MUID:84223957; PMID:6328658

A:Accession: A38022

A:Molecule type: DNA

A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>

A:Cross-references: GB:K02006

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specifi

F:130-395/Domain: protein kinase homology <KIN>

F:138-146/Region: protein kinase ATP-binding motif

F:165/Active site: Lys #status predicted

C;Comment: This sequence is tentative because the introns have not been identified.

C;Genetics:

A;Gene: FlyBase:Egfr

A;Cross-references: FlyBase:FBgn0003731

A;Map position: 2 57F

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-732/Domain: extracellular #status predicted <EXT>

F;733-764/Domain: transmembrane #status predicted <TM>

F;765-1330/Domain: intracellular #status predicted <INT>

F;808-1072/Domain: protein kinase homology <KIN>

F;816-824/Region: protein kinase ATP-binding motif

F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #status p

F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F;843/Active site: Lys #status predicted

F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predic

Query Match 24.2%; Score 1645.5; DB 1; Length 1330;

Best Local Similarity 29.8%; Pred. No. 3.9e-60;

Matches 411; Conservative 181; Mismatches 416; Indels 371; Gaps 39;

QY 80 VQGYVLIHQVROVPLQRLIRIVRGTLGF-----EONVALAVLDNGDPLNNTTPVTGASP 134

DB 38 ITNYIVIGLDLIFCTLSYRLQIIRGTRFLSLSVEEEKYALFV-----TY 81

QY 135 GGLRELQRLSLTBILKGVLIQORNPOLCYODTILMKDQYIKANSKFIGITELNRSRACHP 194

DB 82 SKMYTLEIPDLRLVNGQVGFHNNYLNCHMRTTQWSEIVSNGTDVYNYDFTAPERECPK 141

QY 195 CSPMCKGSRGWGESSDCQSLRTVTCAGGCA--RCKGPLPTDCHCQCAAGCTGPKHSDC 252

DB 142 CHESCTHG--CWGEGPKNCKFKSLTCSPOCAGRCYGPCKPRECCHLFCAGGCTGPTQKDC 200

QY 253 LACLHFNHSGICELHCPALVYNTDTFESNPENPEGRYTFGASCVTACPNYLSLTDVGSCT 312

DB 201 IACKNFFDEAVSKEECPMPKRYNTTYVLETNPEGYAYGATCVKECP--GHLLRDNGACV 259

QY 313 LVCPLHNQVETADGTQRECKSKPCARVCYGLMGHEHLREVRATVSANIQEFAGCKKIFG 372

DB 260 RSCPDQKMDKGGE-----CVPNGPCPKTCPGVTVLH-----AGNIDSPNCTVIDG 306

QY 373 SLAFLPESFDG--DPASNTA-----PLOEQLOVFETLEEITGYLIYISAWPDSLPLDSV 424

DB 307 NTRILDQTSFGQDVVANTMTGPRYIPLOPERREVEFSTVKEITGYLNIETHPTQPNRLSY 366

QY 425 FQNLQVIRGRIHLNGAY--SLTLOGLGISWGLRSLRELGLALIHNTLHCFVHTVPWD 483

DB 367 FRNLETIHGQLMESMFAALAIKSSLSLEMRNLKQISSGVVIOHNRDLCTVSNIRWP 426

QY 484 QLFNRPHQALLHTANRPEDEC-----504

DB 427 ATOKEPEQKQVWYNENLRADLCCKFLTLISVQHNIIMHIFAICREKWNHLLGSGVQRGLL 486

QY 505 -----504

DB 487 GSWHGSVPYLOQLQFQWHLHRLWLVIQVINSTQKSNHQLTDACYSPSVPTSLTIER 546

QY 505 -----VSGELA-- 510

DB 547 ARYAIQSAGLAMELEQITARSASMRHSTKLPAGRQVPRVWFLGVCSARAGIAEPLAGR 606

QY 511 -----CHQLCARGHCWGPCTQVNCQFLRQECVCEVRVLOGLPREYV---NARHCLP 562

DB 607 AVCRKCHPLCELCTNYGHEQVCSTHYKREQCETEC-----PADHYTDEEQRECFQ 660

QY 563 CHPECQPNQSGVTCFPGPADQCVACAHYK-----DPPF-----CVARCPGKYK-PDL 608

DB 661 RHPEC---NG---CTGPGADCKSCNFKLPANETGYPVNSTMPNCTSKCPLEMRHNY 714

QY 609 SYMPIWKFPDEBAGACQPCPINCTHSCVDLDKGCAPAEQORASPLTSIVSAVVGILLVAVLG 668

DB 715 QVTAIGPY-----CAASPPRRSKITANLD-----VNMIPIITGAVLVPTIC 755

QY 669 VVFGI-LIKRQOKIRKYT--MRRLLQETELVPLTPSGAMPNQAOQMRILKETELRKVKV 725

DB 756 ILCVVYIICQKQAKKETVMTWALSGREDSPLPSNIGANLCKURIKADAEURKGV 815

QY 726 LGSAGFTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEIILDEAYVMAGVGVSPVSR 785

DB 816 LGMGAFGRVYKGVWPEGENVKIPVAIKELLKSTGAESSEFLREAYIMASEEHVNLKL 875

QY 786 LGICLSTVQLVTQMLPYGCLLDHVRENRRGLSQDLNWKCMQIAKMSYLEVRLVHRD 845

DB 876 LAVCMSSQMLITQMLPGLCLLDYVRNRDKIGSKALLNMSTQIAKMSYLEERLVRD 935

QY 846 LAARNVLVK---SPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILLRRRTHQ 902

DB 936 LAARNVLRLLAGEDH----DFGLAKLLSSDSNEYKAAGGMPIKWLALCEICINRVFTSK 991

QY 903 SDVMSGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPICITDVTVMVVKCMID 962

DB 992 SDVMAFGVTIWELLTFQRPHEINIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLD 1051

QY 963 SECRPFRELVSEFSRMDPQPFVVIQNEIDLG--PASPLDSTFFYSRLLEDD---DMGDL 1017

DB 1052 AAMRPTFKQTTVFAEFARDPGRYLAIGDKFTRLPA-----YTSQDEKDLIRKLAPT 1104

QY 1018 VDAEYELVPQGFPCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEEEAP----- 1071

DB 1105 TDGSEAIKPDYDYLQPKAALGPS-----HRTDCT-----DEMPKLNRYC 1143

QY 1072 RSLAPSEGAGSDVFDG---DLGMAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYV 1128

DB 1144 KDPNSKNSTGDDERDSSAREVGNGLR-----LDLPVDEDDYL 1182

QY 1129 APLTCSPOEYVQPDVRRPQPSREGPLPAARPAAGATLERAKTLSPGXNGVVKDVFAFG 1188

DB 1183 MP-TCOPGNNNNNM-----NPNQNNMAAVGAAGYM-----DLIGVP 1220

QY 1189 GAVENPEYL---TPGGAAPOPH-----PPPAFSP-AEDNLVYWD 1224

DB 1221 VSDNPEYLLNAQTILGVGESPIQTQIGIPVGGPGTMEVKVMPGSEPTSSDHEYND 1279

RESULT 15

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus

C;Species: avian erythroblastosis virus

C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997

C;Accession: S00727

R;Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.

Oncogene Res. 1, 265-278, 1987

A;Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mut

A;Reference number: S00727; MUID:88217326; PMID:2897102

A;Accession: S00727

A;Molecule type: DNA

A;Residues: 1-545 <SCO>

A;Cross-references: EMBL:X06943

C;Genetics:

A;Gene: erbB

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; phosphotransferase

F;135-400/Domain: protein kinase homology <KIN>

F;143-151/Region: protein kinase ATP-binding motif

Query Match

Best Local Similarity 24.1%; Score 1640; DB 2; Length 545;

Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPFPCVACRCPGKPDLSYMPKFPDEEGACQPCPINCTHSCVDL 637

DB 1 GP--DHCMKCAHFTDGHPCVACFAGVLGENDTL-VWKYADANAVCQCHPNCNTRGCKGP 57

QY 638 DDKCPAEQARASPLTSIVSAV--GILLVVLGVVFGILIKRQOKIRKYTWRRLLQETEL 696

DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVVGLGILYLR--HIVRKTLRRLQLEREL 113

```
QY 697 VEPLTPSGAMPNOAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNOAHLRIKETEFKVKVLGFGAGFTVYKGLWIPEGEKVITIPVAIKEL 173
QY 757 ENTSPKANKIILDEAYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCCLLDHVRENKR 816
Db 174 EATSPKANKIILDEAYVMASVDNPHVCRLGICLTSTVOLITOLMPYGCCLLDYIREHKN 233
QY 817 LGSODLLNMCQIAKGNLYLEDVRLVHRDLAARNVLKSPNVKXITDFGLARLLDIDETE 876
Db 234 IGSYLLNMCVQIAKGNLYLEERHLVHRDLAARNVLKTPQDKITDFGLAKQLGADEKE 293
QY 877 YHADGGKVPKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLE 936
Db 294 YHAEKKVPKWMALLESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 353
QY 937 KGERLPQPPICTIDVYIMVKWMIDSECRPRELYSEFSRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICTIDVYIMVKWMSDADSRPKFRELIAEFSKWARDPPRYLVIOGDERMH 413
QY 996 PASPLDSTFYRSLLEDDMGDLVDAEYLYPQOGFFCPDPAPGAGGMVHRRHSSSTRSG 1055
Db 414 LPSFTDSKFYRTLMEEDMEDIVDAEYLYPHQGF-----NSPST--- 454
QY 1056 GBDLTGLPSEBEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLO 1110
Db 455 -----SRTPLLSLSATSNSATNCIDRNG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNOQDVRPQPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKPSAMVQNIYNYISLTAISK 523
QY 1170 AKTLPCKNGWKVDVFAFGGAVENPEYL 1197
Db 524 LPMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:08:35
Job time : 30.9062 secs